Phase-Normalized Neural Network for Linearization of RF Power Amplifiers

Arne Fischer-Bühner, *Graduate Student Member, IEEE*, Lauri Anttila, *Member, IEEE*,
Manil Dev Gomony, *Member, IEEE*, and Mikko Valkama, *Fellow, IEEE*

Abstract— This letter proposes a methodology for phase-normalization of the complex-valued I/Q inputs of a real-valued time delay neural network (RVTDNN). The normalization enables modeling of the nonlinear behavior of a radio frequency (RF) power amplifier (PA) in a more efficient way, by complying with the physical characteristics of the distortions at RF. The presented digital predistortion (DPD) linearization experiments with a Doherty GaN PA at 3.5 GHz show a 4-dB improvement in the output linearity compared to state-of-the-art neural network (NN) and polynomial-based DPD models, allowing linearization to below −50 dBc adjacent channel leakage ratio (ACLR) levels with feasible processing complexity.

Index Terms— Behavioral modeling, digital predistortion (DPD), nonlinear distortion, radio frequency (RF) power amplifier (PA), time-delay neural network (NN).

I. INTRODUCTION

HIGH transmission linearity is key to achieving the data rates envisioned for 5G-advanced and 6G communications. Yet, the trend toward higher bandwidth and high peak-to-average power ratio (PAPR) of modern transmit waveforms challenge linear amplification at high efficiency. Efficient power amplifiers (PAs), supporting such high bandwidth, demand additional correction to ensure linear transmission. Digital predistortion (DPD) is the de facto implementation approach for linearizing PAs, wherein a distortion model is applied at baseband to cancel out the PA's nonlinear and dynamic passband behavior [1]. Several Volterra- and polynomial-based models, such as the generalized memory polynomial (GMP) [2], or the dynamic deviation reduction (DDR) model [3] have been successfully applied. However, these models face limitations when scaled to greater accuracy, as the high correlation of polynomial kernels impedes the identification of suitable model coefficients [4].

As high linearity is a crucial requirement for high throughput communication, neural networks (NNs) are being researched as a potential modeling solution, offering enhanced scalability and generality compared to polynomial approaches. The real-valued time-delay NN (RVTDNN) is proposed for

PA modeling and DPD in [5], [6] where the in-phase (I) and quadrature-phase (Q) signal parts are individually provided to a single time-delay NN. Since the nonlinear distortions at radio frequency (RF) mainly depend on the envelope of the passband signal, it is suggested in [6] to augment the RVTDNN by providing the baseband amplitudes as NN inputs, as well as powers thereof. Deep, convolutional, and residual feed-forward NN structures are discussed in [4], [7], [8], and [9], which all rely on a similar input data configuration with decomposed I and Q inputs, and the same approach has been employed for linearizing a load-modulated balanced PA [10], beamforming, or MIMO transmitters [11], [12], [13], for joint DPD and PAPR reduction [14], or self-interference cancellation in full-duplex radio [15], [16]. Recurrent NN (RNN) structures have also been studied as an alternative to feed-forward NNs, suited especially for strong PA memory effects [17], [18], [19], [20]. RNNs are, however, complex to train, since the recurrent structures need to be unrolled to ensure temporal consistency during training. This work thus limits itself to feed-forward NN structures.

Although RVTDNN can be scaled to outperform traditional approaches, it can be challenged whether a real-valued NN, provided with separated I/Q baseband signals, can efficiently model the RF behavior of the PA, since the real-valued nature of the NN prevents it from understanding the role of the complex baseband phase for the bandpass nonlinearity. Instead, the RVTDNN will model the distortion in a 2-D I/Q space and consequently needs to learn the nonlinear mapping for each individual I/Q phase angle, without support for the physical nature of the problem. On the other hand, the high generality of the RVTDNN allows modeling effects that are nonphysical at RF, including I/Q imbalance [6]. This generality comes at the cost of a high complexity. The problem is addressed in [21] with a physically inspired model, using decomposed I/Q phase differences alongside the signal envelope to produce an absolute output and a phase difference. While complying with the physical origin of the distortions, the NN has to map inputs from the I/Q space to an angular phase difference, which complicates the modeling. In [22], the vector-decomposed time-delay NN (VDTDNN) is proposed, where only envelope inputs are provided to a fully connected NN layer. The NN is followed by a linear structure, to recover the complex phase. This phase recovery is compatible with the bandpass characteristic of the distortion; yet, it allows only for a linear dependency on past phases. The phase, however, is a nonnegligible component, crucial for the modeling of PM–AM and PM–PM distortions [21]. Thus, the VDTDNN has limited capability due to a lack of relevant degrees of freedom, which is why an RVTDNN will outperform the VDTDNN, when provided with sufficiently many nodes.
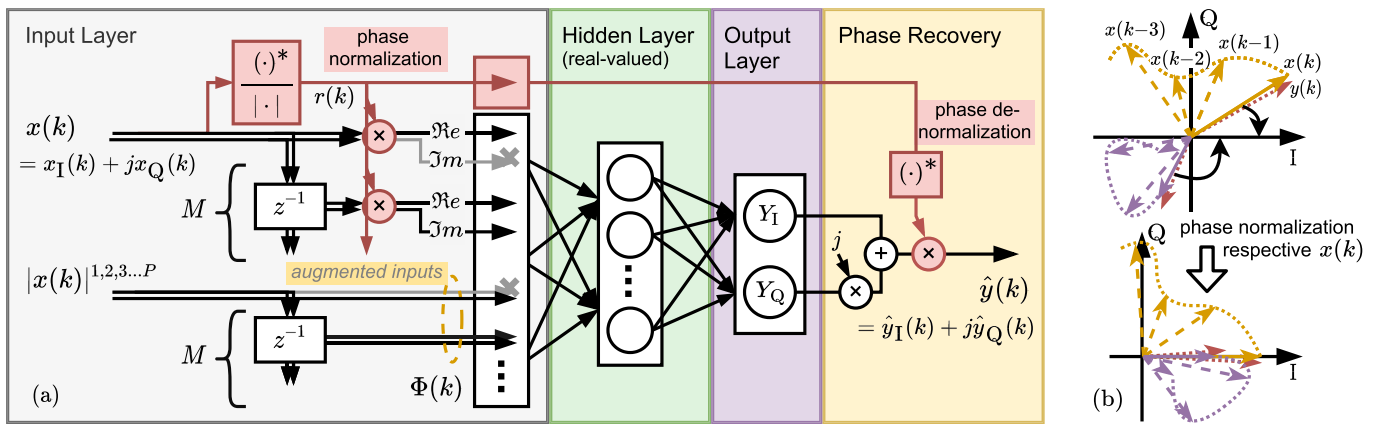
Fig. 1. (a) RVTDNN with proposed phase normalization, highlighted in red. The phase of the I/Q inputs is normalized by applying $r(k)$ and restored at the NN output. Grayed inputs are additionally pruned as redundant or zero. (b) Illustration of phase normalization for two example inputs (yellow and purple).

This letter presents a novel approach for RVTDNN models with decomposed I/Q inputs, which yields a significant improvement in modeling accuracy. The proposed method is compliant with the physical characteristics of the nonlinear distortions, allowing the NN to map PA nonlinearity in a more efficient way. Specifically, we propose an I/Q phase normalization for preprocessing the I/Q inputs to an RVTDNN, where the current and past I/Q samples are rotated with respect to the angle of the instantaneous sample. The original phase is added at the output of the network. Thus, the NN will only learn mappings with reference to a normalized phase. The proposed scheme is experimentally evaluated in the context of PA behavioral modeling and DPD linearization, and shown to outperform the state-of-the-art. Section II details the proposed method, while Section III provides the behavioral modeling and DPD measurement results. Section IV concludes the findings of this letter.

II. PROPOSED PHASE NORMALIZATION

During RF up-conversion, the amplitude of complex baseband samples gets mapped to the RF envelope, while the derivative of the baseband phase modulates the frequency of the transmitted waveform. Consequently, the specific baseband phase of the signals is not seen by the nonlinear PA and is irrelevant to the modeling problem. We propose a phase-normalization input layer for preprocessing of the RVTDNN I/Q inputs, which removes the specific phase such that the NN can efficiently focus on memory effects with respect to the relative phases. Thus, instead of modeling the distortion for any possible input phase, the real-valued NN will map the PA behavior with respect to only a single reference phase and provide a complex-valued output relative to that reference, which enables greater accuracy at lower complexity. The specific phase is then restored at the model output. The NN operates on the envelope and the relative phase of the input signal, thus theoretically having unconstrained modeling potential for RF nonlinearity. The proposed scheme is depicted in Fig. 1, with additions to the RVTDNN highlighted in red. As examples, the normalization of two independent sets of NN I/Q inputs is illustrated, where the current samples are rotated toward zero phase and related memory samples are rotated accordingly.

In order to normalize the phase, the I/Q inputs $\mathbf{X}(k)$ are rotated by multiplication with a normalization factor $r(k)$,

such that the instantaneous input has a zero phase. Thus, the normalization is applied on a per processed input sample $x(k)$ and memory I/Q inputs are rotated according to the phase of $x(k)$. The factor $r(k)$ is given by the complex conjugate of the instantaneous input, normalized by its absolute as

$$r(k) = x^*(k)/|x(k)| \quad (1)$$

where x^* denotes the complex conjugate of x. The complex-valued samples coming out of the input delay line of the TDNN are denormalized by multiplying with $r(k)$ as

$$\mathbf{X}(k) = r(k)\,[x(k), x(k-1), x(k-2), \ldots, x(k-M)] \quad (2)$$

and the composite real-valued input vector to the NN reads

$$\mathbf{\Phi}(k) = [\Re(\mathbf{X}(k)), \Im(\mathbf{X}(k)), \mathbf{A}(k), \ldots, \mathbf{A}(k)^P] \quad (3)$$

where $\mathbf{A}(k)$ is a real-valued envelope vector

$$\mathbf{A}(k) = [|x(k)|, |x(k-1)|, \ldots, |x(k-M)|]. \quad (4)$$

In consequence of the phase normalization, the real part of the instantaneous sample in $\mathbf{X}(k)$ will coincide with its absolute value while its imaginary part becomes zero. These zero and redundant inputs can be pruned from the input vector $\mathbf{\Phi}(k)$, additionally reducing the complexity of the NN. We note that any other reference phase angle may be chosen instead of the zero phase without affecting the proposed method.

The input vector $\mathbf{\Phi}(k)$ is processed by a real-valued feedforward nonlinear NN $\Xi[\cdot]$ with an arbitrary architecture, usually comprising one or more hidden-layers with nonlinear activation, followed by a linear output layer. This is formally expressed as

$$\mathbf{Y}(k) = \Xi[\mathbf{\Phi}(k)]. \quad (5)$$

The hidden layers can be fully connected or convolutional layers or residual NN structures [4], [7], [8], [9]. Without loss of generality, NNs with fully connected layers are considered in this letter. The output layer provides two outputs $\mathbf{Y}(k) = [Y_I(k), Y_Q(k)]$, which resemble the I and Q parts, but with reference to the normalized input. For the complex-valued model output the specific phase is restored by multiplying with a denormalization factor given by the normalized instantaneous input, or preferably, the complex conjugate of the already defined normalization factor, expressed as

$$\hat{y}(k) = x(k)/|x(k)|\big(Y_I + jY_Q\big) = r^*(k)\big(Y_I + jY_Q\big). \quad (6)$$
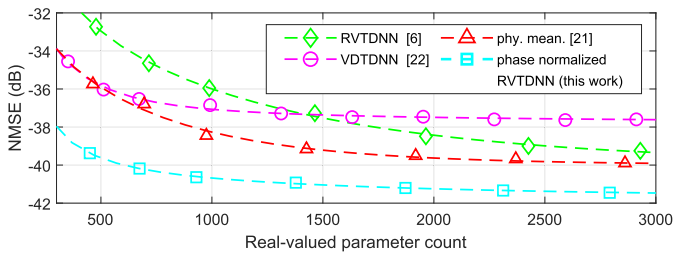
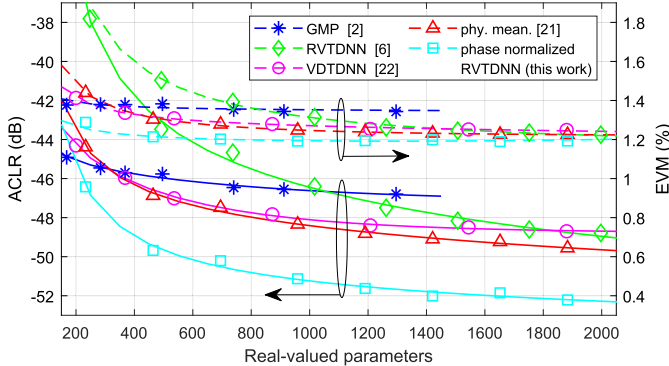Fig. 2. Forward modeling results of the proposed and reference methods.

Fig. 3. Measured ACLR and EVM versus DPD model complexity. EVM performance is bounded by clipper-induced floor of 1.14%.

III. MODELING AND DPD MEASUREMENT RESULTS

First, we demonstrate the proposed method in the context of PA forward modeling. Three models are evaluated with an input-output data of a Doherty GaN PA (*RTH18008S-30*) running a 100-MHz 5G OFDM waveform [30 kHz subcarrier spacing (SCS), 273 resource blocks (RBs)] [23] at 1.8 GHz. The NN models were trained for 1000 epochs on 120k samples using the adaptive moment (ADAM) optimizer, and the normalized mean-squared error (NMSE) was evaluated on 40k samples. Each setting is evaluated five times and results are averaged, to account for the random initialization of the NN coefficients. All the compared models use a memory depth $M = 7$ and first- and third-order envelope terms. The evaluated NN models have one hidden layer with varied neuron count and Sigmoid activation. Fig. 2 shows the achieved NMSE of the RVTDNN with and without the proposed phase normalization, alongside the approaches of [21] and [22]. The modeling results indicate a clear improvement when applying the proposed phase normalization for low as well as high complexity models.

Next, we apply the proposed phase normalization for linearizing a GaN Doherty PA (*QPA3503*) with +35.6 dBm output power, running a 5G compliant 160-MHz multicarrier waveform at 3.5 GHz, with 8.5 dB overall PAPR and four 40-MHz carriers (30 kHz SCS, 106 RBs per carrier) [23]. The *NI PXIe-5840* vector signal transceiver (VST) is used for analog signal generation and RF upconversion, and as receiver for the amplified signal. We utilize the iterative learning control (ILC) scheme [24] to derive an ideal DPD signal, which we then use to train the NN models. The same training configuration is applied as in the forward modeling experiment. Note that the ILC-based training approach was chosen for model evaluation purposes. In an adaptive real-time deployment, ILC based pretraining may be combined with indirect learning [6] or closed-loop adaptation of the models [10]. A GMP [2] predistorter is provided for reference, tuned using least-squares fitting on the ILC data; the complex-valued

TABLE I
DPD EXPERIMENT RESULTS WITH DOHERTY GaN PA AT 3.5 GHz

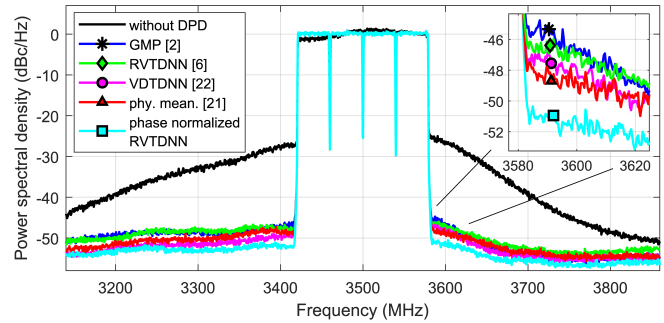| Model	Parameters	Neurons	NMSE	ACLR	EVM
Without DPD	–	–	–17.7 dB	–25.9 dB	6.47 %
ILC baseline	–	–	–49.7 dB	–56.6 dB	1.14 %
GMP [2]	1296	–	–40.3 dB	–46.8 dB	1.34 %
RVTDNN [6]	1226	36	–41.5 dB	–47.5 dB	1.26 %
VDTDNN [22]	1208	56	–42.7 dB	–48.4 dB	1.25 %
Phy. mean. [21]	1190	36	–42.5 dB	–48.8 dB	1.24 %
This work	**959**	**29**	**–44.7 dB**	**–51.1 dB**	**1.19 %**

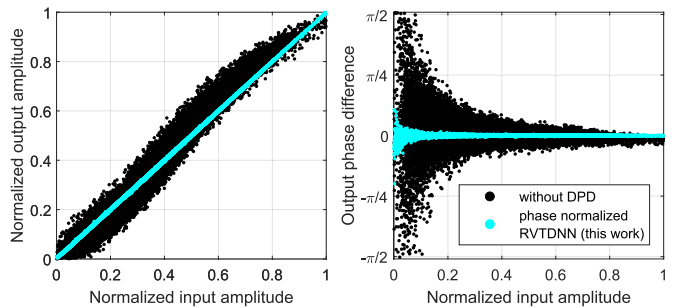Fig. 4. Power spectrum of the PA output using different DPD models.

Fig. 5. AM–AM and AM–PM plots for the original and linearized PA.

coefficients were considered as two real-valued equivalent coefficients for complexity comparison. Fig. 3 depicts the linearization performance in terms of worst adjacent channel leakage ratio (ACLR) and the error vector magnitude (EVM). With the proposed phase normalization applied, the RVTDNN reaches an ACLR of below −50 dBc if scaled to beyond 600 parameters. Table I reports detailed results for the different models and Fig. 4 shows the corresponding power spectra. AM–AM and AM–PM plots showing the linearized behavior are presented in Fig. 5. The proposed phase normalized RVTDNN clearly outperforms the state-of-the-art.

IV. CONCLUSION

A novel phase normalization scheme for NN I/Q inputs was presented, which enables an RVTDNN to efficiently model RF nonlinear distortions at baseband. The provided modeling and DPD linearization RF experiment results show significant performance gains when applying the method, or alternatively, a significant reduction in NN complexity for achieving the same linearization. Furthermore, the phase normalization does not restrict the modeling degrees of freedom and thus allows scaling toward high parameter counts and higher modeling accuracy. The proposed scheme can be combined with any feed-forward-type TDNN model available in the DPD literature.

REFERENCES

[1] J. Wood, "System-level design considerations for digital pre-distortion of wireless base station transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 5, pp. 1880–1890, May 2017.

[2] 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.

[3] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic deviation reduction-based Volterra behavioral modeling of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 12, pp. 4323–4332, Dec. 2006.

[4] R. Hongyo, Y. Egashira, T. M. Hone, and K. Yamaguchi, "Deep neural network-based digital predistorter for Doherty power amplifiers," *IEEE Microw. Wireless Compon. Lett.*, vol. 29, no. 2, pp. 146–148, Feb. 2019.

[5] M. Rawat, K. Rawat, and F. M. Ghannouchi, "Adaptive digital predistortion of wireless power amplifiers/transmitters using dynamic real-valued focused time-delay line neural networks," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 1, pp. 95–104, Jan. 2010.

[6] D. Wang, M. Aziz, M. Helaoui, and F. M. Ghannouchi, "Augmented real-valued time-delay neural network for compensation of distortions and impairments in wireless transmitters," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 242–254, Jan. 2019.

[7] X. Hu et al., "Convolutional neural network for behavioral modeling and predistortion of wideband power amplifiers," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 3923–3937, Aug. 2022.

[8] 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, Mar. 2022.

[9] Y. Wu, U. Gustavsson, A. G. I. Amat, and H. Wymeersch, "Residual neural networks for digital predistortion," in *Proc. IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.

[10] E. Guillena, W. Li, G. Montoro, R. Quaglia, and P. L. Gilabert, "Reconfigurable DPD based on ANNs for wideband load modulated balanced amplifiers under dynamic operation from 1.8 to 2.4 GHz," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 1, pp. 453–465, Jan. 2022.

[11] A. Brihuega, L. Anttila, and M. Valkama, "Neural-network-based digital predistortion for active antenna arrays under load modulation," *IEEE Microw. Wireless Compon. Lett.*, vol. 30, no. 8, pp. 843–846, Aug. 2020.

[12] P. Jaraut, M. Helaoui, W. Chen, M. Rawat, N. Boulejfen, and F. M. Ghannouchi, "Review of the neural network based digital predistortion linearization of multi-band/MIMO transmitters," in *IEEE MTT-S Int. Microw. Symp. Dig.*, May 2021, pp. 1–3.

[13] C. Tarver, A. Balalsoukas-Slimining, C. Studer, and J. R. Cavallaro, "Virtual DPD neural network predistortion for OFDM-based MU-massive MIMO," in *Proc. 55th Asilomar Conf. Signals, Syst., Comput.*, Oct. 2021, pp. 376–380.

[14] Z. Liu, X. Hu, W. Wang, and F. M. Ghannouchi, "A joint PAPR reduction and digital predistortion based on real-valued neural networks for OFDM systems," *IEEE Trans. Broadcast.*, vol. 68, no. 1, pp. 223–231, Mar. 2022.

[15] A. Balatsoukas-Stimming, "Non-linear digital self-interference cancellation for in-band full-duplex radios using neural networks," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2018, pp. 1–5.

[16] Y. Kurzo, A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Hardware implementation of neural self-interference cancellation," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 10, no. 2, pp. 204–216, Jun. 2020.

[17] 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.

[18] T. Kobal, Y. Li, X. Wang, and A. Zhu, "Digital predistortion of RF power amplifiers with phase-gated recurrent neural networks," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 6, pp. 3291–3299, Jun. 2022.

[19] 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, Jun. 2022.

[20] T. Kobal and A. Zhu, "Digital predistortion of RF power amplifiers with decomposed vector rotation-based recurrent neural networks," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 11, pp. 4900–4909, Nov. 2022.

[21] E. G. Lima, T. R. Cunha, and J. C. Pedro, "A physically meaningful neural network behavioral model for wireless transmitters exhibiting PM–AM/PM–PM distortions," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 12, pp. 3512–3521, Dec. 2011.

[22] Y. Zhang, Y. Li, F. Liu, and A. Zhu, "Vector decomposition based time-delay neural network behavioral model for digital predistortion of RF power amplifiers," *IEEE Access*, vol. 7, pp. 91559–91568, Jul. 2019.

[23] *NR; Base Station (BS) Radio Transmission and Reception*, document Tech. Spec. 38.104, V 18.1.0, Release 18, 3GPP, Apr. 2023.

[24] J. Chani-Cahuana, P. N. Landin, C. Fager, and T. Eriksson, "Iterative learning control for RF power amplifier linearization," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 9, pp. 2778–2789, Sep. 2016.