# Augmentation of Self-Interference Cancellation for Full-Duplex using NARX Neural Networks

Qingqing Dong, *Graduate Student Member, IEEE,* Andrew C. M. Austin, *Member, IEEE,* and Kevin W. Sowerby, *Senior Member, IEEE,*

***Abstract*—A self-interference cancellation augmentation technique based on a NARX (Nonlinear Autoregressive Exogenous) network model is implemented and evaluated on an OFDM-based full-duplex system testbed operating at 2.4 GHz. In a comparison with the state-of-the-art polynomial models, our experimental results demonstrate the significant computational efficiency of the proposed NARX model. Specifically, the NARX model with one hidden layer reduces computations by 83.3% while achieving the same cancellation level within a bandwidth of 2 MHz.**

***Index Terms*—In-band full-duplex, NARX model, analog impairments, self-interference cancellation, signal processing**

## I. Introduction

In-band full-duplex radio systems have gained interest due to their potential to double spectrum efficiency by transmitting and receiving a signal simultaneously in the same frequency band [1]. However, a major technical obstacle to realizing this operation is the presence of self-interference. The self-interference signal, originating from the local transmitter, often has higher power levels compared to the desired signal from remote transmitters, leading to reception degradation.

Passive self-interference suppression (increasing the passive isolation between the transmitting and receiving antennas [2]) and analog cancellation (using a replica signal digitally generated with another RF chain and/or tapped by an analog delay line [3], [4]) provide a certain level of self-interference cancellation. However, in general, these alone are insufficient to push the self-interference down to the desired level and are often *augmented* by a digital cancellation stage [5]. Previously, digital augmentation has required the modelling of circuitry imperfections to estimate and remove the distorted self-interference [6], [7], [8], [9]. Our proposed technique does not require explicit knowledge of each analog distortion. Also, implementing the models (e.g., memory polynomials) to capture cascaded effects can be computationally expensive [10], but the NARX model-based technique provides a greater level of cancellation with a significantly lower level of computational cost.

The implementation of machine learning methodologies for mitigating self-interference has shown promising results. For instance, the application of a standard gradient descent algorithm has been explored in self-interference analog cancellation techniques to optimize tuning parameters [3]. Furthermore, it has been demonstrated that the use of machine learning on multi-tap RF cancellers can accelerate the tuning process and significantly improve cancellation performance [11], [12]. In a related work [13], experimental results demonstrate the effectiveness of feed-forward neural networks, which achieved a 36% reduction in computational requirements compared to conventional models, with a 7 dB self-interference nonlinear cancellation (it should be noted that this is *in addition* to passive suppression and analog cancellation).

Contemporary nonlinear cancellation techniques are limited by the need to develop mathematical expressions for various distortion effects and normally require a considerable number of computations and parameters [4], [5], [8]. Our proposed NARX model-based technique can efficiently address the cascaded nonlinear memory effects of power amplifiers [9] and IQ imbalance [7] without requiring the explicit modelling of each nonlinear effect. This approach signifies an additional cancellation gain achieved beyond conventional techniques. The key advantage of this augmented approach lies in its compatibility with other cancellation methods, such as analog and passive techniques, as it operates within the baseband. This feature offers significant flexibility to complement and enhance the performance of existing self-interference mitigation strategies. The primary goal of this paper is to demonstrate the effectiveness of NARX models in mitigating self-interference in full-duplex systems by applying them to an OFDM-based full-duplex experimental system operating at 2.4 GHz. Our research also includes a comparative analysis with a state-of-the-art polynomial model, focusing on technical cancellation performance and computational complexity.

## II. Self-interference Signal Model

The performance of conventional digital cancellation techniques is associated with the accuracy of mathematically modelling the received self-interference signal. Assuming there is a linear and time-invariant channel, the received self-interference signal in this simple case can be expressed as:

$$y_{lin}(n) = \sum_{m=0}^{M-1} h_{SI}(m)x(n-m). \tag{1}$$

Here, $x(n)$ denotes the transmitted digital baseband self-interference signal, which has been sampled at a rate exceeding the Nyquist rate and the bandwidth of the signal of interest. $h_{SI}$ represents the channel impulse response coefficients and $M$ is the maximum delay considered.

Despite the distinct distortions exhibited by each hardware component, self-interference models that account for IQ imbalance and power amplifier distortions have demonstrated

Q. Dong and K. W. Sowerby are with the Department of Electrical, Computer, and Software Engineering, The University of Auckland, Auckland, New Zealand. A. C. M. Austin is with the School of Electrical, Electronic, and Mechanical Engineering, University of Bristol, Bristol, United Kingdom.

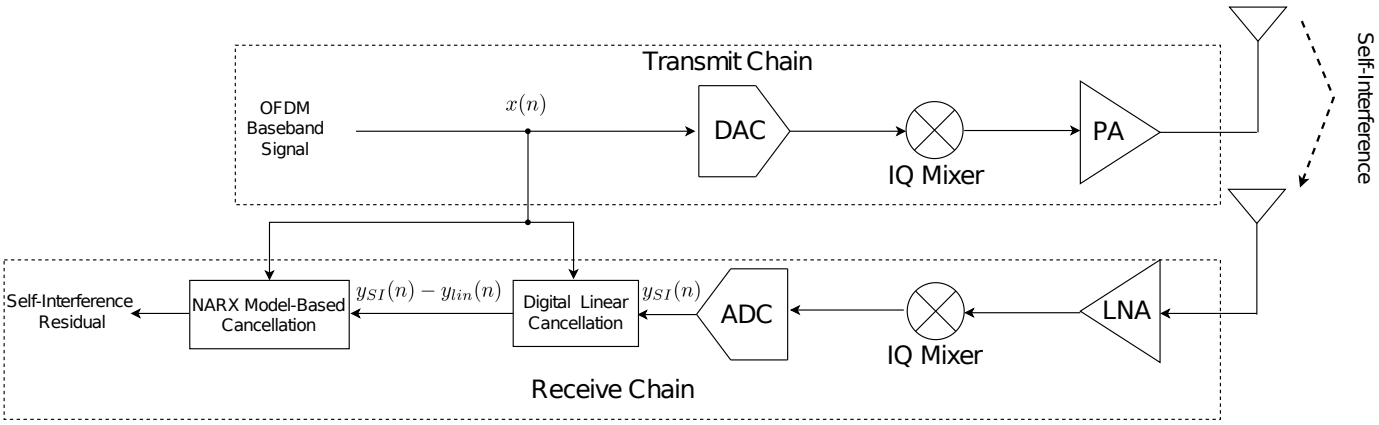


Fig. 1. A block diagram of the proposed in-band full-duplex system architecture with the proposed NARX model-based cancellation augmenting any analog and passive suppression.

effective performance in achieving the desired cancellation level [8]. Accordingly, the expression for the received self-interference signal (1) can be modified to incorporate these effects [8]:

$$y_{SI}(n) = \sum_{\substack{p=1,\\ p\ odd}}^{P} \sum_{q=0}^{p} \sum_{m=0}^{M-1} h_{p,q}(m) x(n-m)^q x^*(n-m)^{p-q}, \tag{2}$$

where $x^*(n)$ denotes the conjugation term of $x(n)$, $h_{p,q}$ denotes the model coefficients, $M$ and $P$ denote the memory depth and the highest nonlinear order number of the model, separately. Generally, the coefficients of the polynomial model are estimated using the normal equation based on the least squares algorithm [9], which involves a lower-upper (LU) decomposition for the matrix inversion [14]. The regenerated self-interference signal based on (2), is then subtracted from the received distorted signal to cancel the self-interference.

However, the number of coefficients which are given by $\left(\left(\frac{P+1}{2}\right)\left(\frac{P+1}{2}+1\right)-1\right)M$ increases exponentially depending on the increase of the highest polynomial order number $P$. As a result, implementing a polynomial model demands a significant computational workload and memory storage. Additionally, the limited adaptability of polynomial models to dynamic environmental changes often requires frequent re-estimation of the model coefficients.

## III. NARX Neural Network Implementation

### A. System Overview

The hardware platform of the experimental in-band full-duplex system mainly consists of a laptop for digital signal processing and an AD9361 module serving as an analog front end. The block diagram of the entire system is shown in Fig. 1. In this setup, a digital QPSK-modulated OFDM signal $x(n)$, comprising 64 subcarriers with a 2 MHz bandwidth, is generated using MATLAB. The signal is then up-converted to a 2.4 GHz RF signal using a DAC (digital-to-analog converter), IQ mixer, and power amplifier. The transmit antenna radiates the signal to the remote receive antenna. Since the local transmit antenna and receive antenna are closely located, the unwanted signal received by the local receiver becomes self-interference. After down-conversion, the distorted signal $y_{SI}(n)$ is digitized. Digital cancellation using a neural network involves two main steps. Firstly, the least squares algorithm is employed to estimate the linear self-interference $y_{lin}(n)$, as expressed in (1), which is then subtracted from the received signal $y_{SI}(n)$. Secondly, the NARX network model is trained using the transmit signal $x(n)$ and the residual self-interference $y_{SI}(n) - y_{lin}(n)$. After training, the NARX model predicts the nonlinear part of the self-interference, which is subtracted from $y_{SI}(n) - y_{lin}(n)$ to obtain the residual self-interference. It is important to note that, for an accurate and repeatable assessment of the NARX model's experimental performance, the antenna separation technique illustrated in Fig. 1 was not utilized. Instead, during the signal power measurement, the transmit and receive antennas were replaced with an attenuator to avoid analog front-end saturation. It is assumed that sufficient analog cancellation has been achieved so that the entire signal can be captured in the dynamic range of the receiver ADC.

### B. Model Structure

After the digital linear cancellation subtracts the major linearities, the NARX model, as an alternative to conventional models, is designed to capture the nonlinear patterns induced by the hardware. Fig. 2 shows the structure of the implemented NARX network model with one hidden layer and one output node during the prediction phase. Each circle on the diagram represents one node, and the weights represented by the arrows interconnect those nodes between different layers. As neural networks can only handle real numbers, the complex transmitted self-interference baseband signal $x(n)$ is divided into real and imaginary parts $\Re\{x(n)\}$ and $\Im\{x(n)\}$. Through time-delay lines, where one sample instant is represented by $z^{-1}$, the input time-series data $\Re\{x(n)\}$ and $\Im\{x(n)\}$ are converted into the input datasets.

Two NARX models with the same parameter settings are used to predict the real and imaginary parts of the distorted signal $y(n)$ separately, but the inputs for those neural networks are the same. This is for modelling the IQ imbalance more

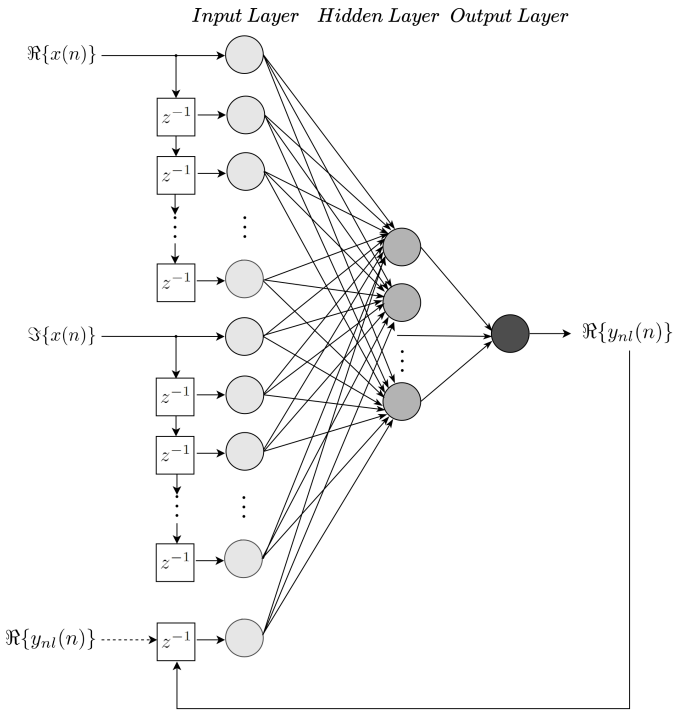


Fig. 2. NARX neural network structure with one hidden layer and one output during prediction phase.

accurately, which results in different influences on in-phase and quadrature components. The predicted output is fed back to the input node via a time-delay line, thus forming a closed-loop structure shown in the diagram for predicting sequential data of historical dependencies. It is also noted, during the training phase, the information of the received self-interference signal is directly provided to the model before prediction.

### C. Model Hyperparameters

Hyperparameters play a significant role in determining the architecture and behaviour of the neural network. They are not learned from the data but are set before training. Different choices of hyperparameters can lead to variations in the network's ability to learn, generalize, and converge to an optimal solution. In this paper, we consider the impact of the structural complexity of the NARX model on cancellation. Models with different numbers of hidden layers were evaluated. Model $A$, featuring a simple architecture with a single hidden layer, was effectively trained with a 20% subset of the total dataset. In contrast, Model $B$ adopted a more complex structure with two hidden layers, demanding a larger training dataset. To ensure the optimal training of Model $B$, a 40% training percentage was employed.

The other hyperparameters for the experiment were carefully determined through a comprehensive process of trying out various combinations and values. The batch size corresponds to one OFDM symbol length of 80, allowing the processing of a complete symbol each time. The epoch size of 100 specifies how many times the entire dataset is processed. These choices strike a balance between training speed and results, considering the dataset's characteristics and model architecture. The regularization parameter was assigned a value of 0.0001. The activation function used at the hidden layer was the $\tanh(\cdot)$ function. The cost function to measure the network learning performance was the mean squared error function, and the optimization algorithm employed was the Levenberg-Marquardt backpropagation algorithm. The weights and biases were generated using the Nguyen-Widrow initialization algorithm, which ensures an even distribution of the active region of each neuron and contributes to accelerating the training progress of the backpropagation algorithm.

## IV. Experimental Results

### A. Cancellation Performance

Fig. 3 depicts the power spectrum of the transmitted signal, the signal after digital linear cancellation, the residual signal after applying the polynomial model and the NARX models, and the noise floor. The signal power values were calculated using the Welch's method, a widely used power spectral density estimation technique. The noise floor was measured at the receiver when transmitting all zeros, and it represents the achievable level of cancellation techniques. The digital linear cancellation effectively reduces the power of major linearities by approximately 20 dB. Using the NARX and polynomial models could both achieve a further decrease in the residual self-interference. The shallow NARX network (Model $A$) outperforms the polynomial model by 0.7 dB, while the deeper network (Model $B$) demonstrates an improvement of approximately 3 dB. This is consistent with the hypothesis that a deeper network has the potential to learn and represent the more complex patterns in data. It is also observed that the polynomial model-based cancellation is not that efficient in eliminating harmonic distortions and LO leakage, whereas these limitations are addressed by the NARX Model $A$ and significantly improved by the NARX Model $B$. This difference could be attributed to the inadequate consideration of these impairments in polynomial models, while the NARX models explicitly represent them.

### B. Comparison of Computations

The required FLOPs (floating-point operations) for achieving nonlinear cancellation performance with polynomial and NARX models are shown in Fig. 4. It is noted that the self-interference cancellation levels specified in this plot are considered cancellation gain, which focuses on the cancellation of the nonlinearities and enhances the practical performance after the digital linear cancellation is implemented, as Fig. 1 implied. Increasing computations (the order number and memory depth) generally lead to an improved cancellation performance, but the NARX models exhibit a more consistent growth in calculations as cancellation results improve compared to the polynomial model. Notably, NARX Model $A$ and $B$ require significantly fewer computational resources than the polynomial model at all cancellation levels. For example, at approximately 9 dB cancellation, NARX Model $A$ with 32 input nodes and one hidden layer of two nodes requires only 16.7% of the FLOPs needed by a 7th-order polynomial model. Additionally, Model $B$ augment the nonlinear cancellation by

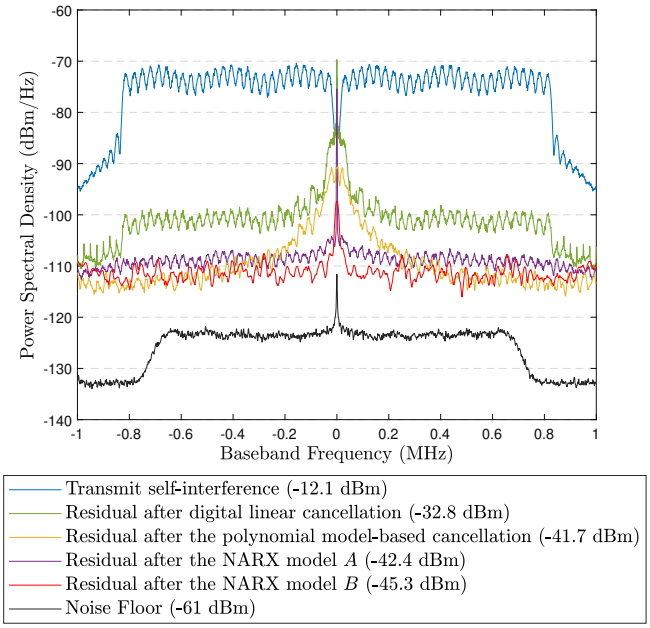


Fig. 3. Power spectral densities of the self-interference signal, the self-interference signal after digital linear cancellation, the residual signal after the 7th-order polynomial model-based cancellation with $M = 32$, the residual signal after the NARX model-based cancellation, and the noise floor.

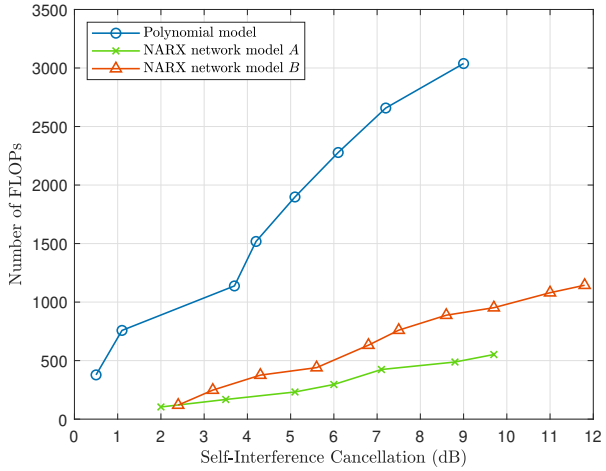


Fig. 4. Comparison of computational complexity between the NARX and polynomial models for self-interference nonlinear cancellation in full-duplex systems.

approximately 3 dB while reducing computations by 62% compared to the polynomial model. However, it should be noted that due to the structural complexity, Model $B$ requires nearly twice the computation of Model $A$ to achieve the same cancellation performance. Improving the neural network performance is mainly done by adjusting the number of input nodes, but significant improvement generally requires the increase of hidden layer size.

## V. Conclusions

In this paper, we propose a novel self-interference cancellation technique based on the NARX model and evaluate its performance through experiments on an in-band full-duplex experimental system. This digital-domain cancellation augments any passive suppression or analog cancellation in a full-duplex system. The results demonstrate that the NARX model outperforms the conventional polynomial model in modelling analog nonlinearities while requiring fewer computations and parameters. Furthermore, our investigation of NARX models with different structures reveals that using deeper networks can enhance cancellation performance. However, it is important to carefully select the depth of the neural network to strike a balance between the desired performance and computational complexity.

## References

[1] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Sep. 2014.

[2] V. Panse, T. K. Jain, P. K. Sharma, and A. Kothari, "Digital Self-interference cancellation in the era of machine learning: A comprehensive review," *Physical Commun.*, vol. 50, pp. 101526, 2022.

[3] K. E. Kolodziej, A. U. Cookson, and B. T. Perry, "Adaptive learning rate tuning algorithm for RF self-interference cancellation," *IEEE Trans. Microwave Theory Tech.*, vol. 69, no. 3, pp. 1740–1751, Mar. 2021.

[4] S. Ge, J. Meng, J. Xing, Y. Liu, and C. Gou, "A digital-domain controlled nonlinear RF interference cancellation scheme for co-site wideband radios," *IEEE Trans. Electromagn. Compat.*, vol. 61, no. 5, pp. 1647–1654, Oct. 2019.

[5] E. Ahmed, A. M. Eltawil, and A. Sabharwal, "Self-interference cancellation with nonlinear distortion suppression for full-duplex systems," in *Proc. Asilomar Conf. Signals, Syst., Computers*, pp. 1199–1203, 2013.

[6] A. Balatsoukas-Stimming, P. Belanovic, K. Alexandris, and A. Burg, "On self-interference suppression methods for low-complexity full-duplex MIMO," in *Proc. Asilomar Conf. Signals, Syst., Computers*, pp. 992–997, Nov. 2013.

[7] G. Fettweis, M. Lohning, D. Petrovic, M. Windisch, P. Zillmann and W. Rave, "Dirty RF: a new paradigm", in *Proc. IEEE 16th Int. Symp. Pers. Indoor Mobile Radio Commun.*, vol. 4, pp. 2347-2355, 2005.

[8] D. Korpi, L. Anttila, and M. Valkama, "Nonlinear self-interference cancellation in MIMO full-duplex transceivers under crosstalk," *EURASIP J. Wirel. Commun. Netw.*, vol. 2017, no. 1, p. 24, 2017.

[9] L. Ding and G.T. Zhou and D. R. Morgan and Z. Ma and J. S. Kenney and J. Kim and C. R. Giardina, "A robust digital baseband predistorter constructed using memory polynomials", *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 159-165, Jan. 2004.

[10] A. Balatsoukas-Stimming, A. C. M. Austin, P. Belanovic, and A. P. Burg, "Baseband and RF hardware impairments in full-duplex wireless systems: experimental characterization and suppression," *EURASIP J. Wireless Commun. Netw.*, vol. 2015, no. 1, p. 142, May 2015.

[11] K. E. Kolodziej, A. U. Cookson and B. T. Rerry, "Machine learning for accelerated IBFD tuning in 5G flexible duplex networks", in *Proc. IEEE/MTT-S Int. Microw. Symp.*, pp. 691-694, 2020.

[12] H. Guo, J. Xu, S. Zhu and S. Wu, "Realtime software defined self-interference cancellation based on machine learning for in-band full duplex wireless communications", in *Proc. Int. Conf. Comput. Netw. Commun.*, pp. 779-783, 2018.

[13] 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)*, pp. 1-5, Jun. 2018.

[14] MathWorks. (2023). *Matrix inverse* [Online]. Available: https://au.mathworks.com/help/matlab/ref/inv.html