On Deep Learning Assisted Self-Interference Estimation in a Full-Duplex Relay Link

Konstantin Muranov, Md Atiqul Islam, *Graduate Student Member, IEEE*, Besma Smida, *Senior Member, IEEE*, and Natasha Devroye, *Senior Member, IEEE*

***Abstract*—We propose and evaluate two new methods for digital mitigation of the non-linear Self-Interference (SI) signal in the context of a Full-Duplex (FD) relay link. The proposed methods utilize Deep Neural Networks (DNN) for the reconstruction of the non-linear SI components. Existing DNN-based SI mitigation techniques require a large number of on-line training samples, which takes resources away from actual data communication. The proposed methods allow a significant reduction in or complete elimination of on-line training by utilizing transfer learning and decoupling the reconstruction of non-linear transmitted SI signal from the estimation of the SI propagation channel, respectively. We utilize link-level simulation to demonstrate the SI cancellation performance of the proposed approaches compared to the existing state-of-the-art DNN-based solutions. The simulation results confirm that the proposed SI mitigation techniques achieve robust Bit Error Rate (BER) performance for Signal-to-Interference-Ratio (SIR) levels as low as −20dB. We derive an upper bound and an approximation for the probability of generalization error that is caused by an insufficient training set size.**

***Index Terms*—Full-duplex relay, deep neural networks, non-linear self-interference, transfer learning.**

I. Introduction

FULL-DUPLEX technology has attracted a significant amount of research due to its potential spectral efficiency gains through the reuse of the same time and spectral resources for uplink and downlink communication [1]. Two-hop FD relay systems are of special interest since they can be seamlessly inserted when a direct link between the source and destination nodes is not viable. Performance of an FD transceiver depends on the mitigation of SI signal stemming from simultaneous transmission and reception. SI cancellation is accomplished by a combination of physical antenna separation as well as analog and digital cancellation [2], [3]. A key challenge in SI mitigation is dealing with non-linear distortion introduced at the transmitter. There are a number of well-known non-DNN signal processing techniques for estimating the SI affected by non-linear distortion, [4]–[7]. These traditional methods require prior knowledge of the non-linear model such as polynomial order. In contrast, the DNN-based solutions [8]–[11], do not require prior knowledge of the non-linear distortion model. Also, it has been shown in [8] that utilization of DNN can reduce the computational complexity by as much as 36% relative to a non-DNN-based solution [4]. Existing DNN-based methods combine estimation of the mostly time-invariant non-linear distortion and the time-varying SI propagation channel, which needs to be periodically reacquired. Hence, the need for *on-line* training. Also, these methods do not leverage the information available from the previous time intervals. This letter aims to improve the spectral efficiency by minimizing the amount of *on-line* DNN training, since it causes interruption of the data transmission. This is accomplished by decoupling the time-varying SI propagation channel from the estimation of time-invariant non-linear distortion or by utilizing *transfer learning*, which combines elements of both off-line and on-line training and accumulates the information across the time intervals.

Manuscript received June 7, 2021; revised July 23, 2021; accepted September 22, 2021. Date of publication September 29, 2021; date of current version December 9, 2021. The work of Besma Smida and Md Atiqul Islam was supported in part by NSF under Award 1620902. The work of Natasha Devroye was supported in part by NSF under Award 1815428. The associate editor coordinating the review of this article and approving it for publication was Q. Wu. *(Corresponding author: Konstantin Muranov.)*

The authors are with the Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL 60607 USA (e-mail: kmuran2@uic.edu; mislam23@uic.edu; smida@uic.edu; devroye@uic.edu).

Digital Object Identifier 10.1109/LWC.2021.3116188

A. Contributions

We propose and evaluate two new DNN-based non-linear SI cancellation methods that significantly reduce or eliminate the on-line training time while delivering SI reduction on par with or better than [4], [8]. The proposed techniques perform well under time-varying SI propagation channel.

- *TID Method (Time-Invariant non-linear Distortion):* In contrast to [8], [9], [11], DNN is only used for estimation of the time-invariant non-linear distortion, while traditional Least-Squares-based method is used for estimation of the time-varying SI propagation channel. This eliminates the need for on-line training assuming that non-linear distortion does not change over time due to aging.
- *TVD Method (Time-Varying non-linear Distortion):* The architecture is similar to [8] and [11], but the amount of on-line training is significantly reduced by using transfer-learning-based DNN training. This method supports estimation of time-varying non-linear distortion.

The proposed SI cancellation methods are considered in the context of the two-hop FD relay system of Fig. 1. The proposed techniques achieve 2–3dB of SIR gain and up to 1-dB BER performance gain (at BER $= 10^{-4}$) compared to [4] and [8]. To the best of the authors' knowledge, this is the first time the end-to-end performance of such relay system employing DNN-based SI cancellation has been evaluated. Finally, we derive an upper bound (8) and an approximation (18) for the generalization error probability. This approximation can guide the selection of training set size as confirmed by simulation.

II. System Model and Notation

A. System Model

The system of Fig. 1 consists of a Half-Duplex (HD) source, a FD relay node, and a HD destination node. The notation

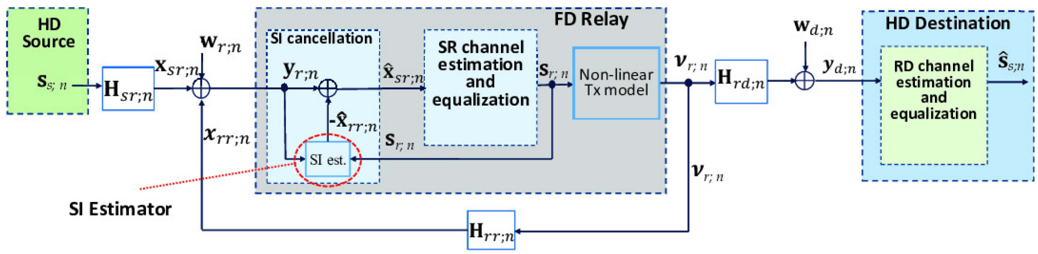

Fig. 1. Two-hop FD relay link employing DNN-based SI cancellation at the FD relay node receiver.

TABLE I
NOTATION SUMMARY

notation	description of the quantity
$\mathbf{s}_{s;n}$	signal vector transmitted by the source.
$\mathbf{s}_{r;n}$	relay Tx signal before non-linear distortion.
$\boldsymbol{\nu}_{r;n}$	relay Tx signal after non-linear distortion.
$\mathbf{h}_{sr;n}$,$\mathbf{h}_{rd;n}$,$\mathbf{h}_{rr;n}$	source-relay (SR), relay-destination (RD), and self-interference (SI) channels, respectively.
$\mathbf{w}_{r;n}$, $\mathbf{w}_{d;n}$	AWGN at the relay and destination, respectively.
$\mathbf{x}_{sr;n}$	signal at the output of SR channel.
$\mathbf{x}_{rr;n}$	signal at the output of SI channel.
$\mathbf{x}_{rd;n}$	signal at the output of RD channel.
$\mathbf{y}_{r;n}$	combined RX signal and noise at the relay.
$\mathbf{y}_{d;n}$	combined RX signal and noise at the destination.
M	channel order.
L	number of sub-channels (oversampling rate).
i	sampling phase, $i = 0, ..., L-1$.
A	data packet size ($A >> M$).
p	delay (in symbol intervals) in relay transceiver.
n	data packet index.

is summarized in Table I. The system employs fractionally-spaced channel equalization with oversampling rate L. The data packets of size A symbols are transmitted by the source. The source-relay (SR), relay-destination (RD), and SI propagation channels are assumed to be frequency-selective with realizations that are statistically independent for each data packet. The spectral efficiency is improved by utilizing blind SR and RD channel estimation at the relay and destination, respectively. The signal at the relay receiver is defined as

$$\mathbf{y}_{r;n}^{(i)} = \mathbf{H}_{sr;n}^{(i)}\mathbf{s}_{s;n} + \mathbf{H}_{rr;n}^{(i)}\boldsymbol{\nu}_{r;n} + \mathbf{w}_{r;n}^{(i)}. \tag{1}$$

The signal at the destination node receiver is given by

$$\mathbf{y}_{d;n}^{(i)} = \mathbf{H}_{rd;n}^{(i)}\boldsymbol{\nu}_{r;n} + \mathbf{w}_{d;n}^{(i)}, \tag{2}$$

where realizations of the sub-channel impulse responses are assumed to be statistically independent from each other. The AWGN components for each sub-channel are also assumed to be statistically-independent. A more detailed system description can be found in [12] and [13]. Using the parallel Hammerstein (PH) model [4] for the non-linear signal transmitted by the relay, the latter can be expressed at sample m as

$$\nu_{r;n}(m) = \sum_{j=1,odd}^{J}\sum_{t=0}^{T} h_{nl}(t)s_{c;n}(m-t)|s_{c;n}(m-t)|^{j-1}, \tag{3}$$

where j is the non-linear term order, the impulse response coefficients of PH model are given by $h_{nl}(t)$, and $s_{c;n}(m)$ models the IQ imbalance introduced during the carrier modulation,

$$s_{c;n}(m) = a_1 s_{r;n}(m) + a_2 s_{r;n}^{*}(m). \tag{4}$$

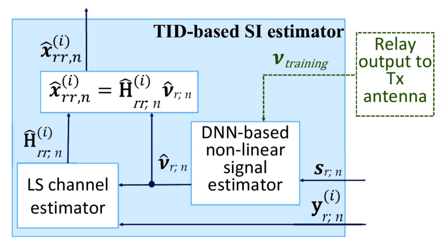

Fig. 2. DNN-assisted SI estimation TID method.

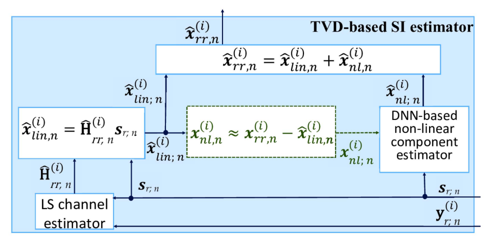

Fig. 3. DNN-assisted SI estimation TVD method.

We use a simplified cubic polynomial version of (3) from [14].

B. Problem Statement

We propose and evaluate two DNN-based methods for estimating the SI defined as,

$$\mathbf{x}_{rr;n}^{(i)} = \mathbf{H}_{rr;n}^{(i)}\boldsymbol{\nu}_{r;n}, \tag{5}$$

where $\boldsymbol{\nu}_{r;n}$ is the transmitted signal affected by non-linear distortion. Our aim is to estimate the received SI signal, $\mathbf{x}_{rr}$ based on the pre-distortion transmitted sequence, $\mathbf{s}_r$ in (4), and the received signal $\mathbf{y}_r$. We use both DNN and traditional signal-processing techniques, please see survey [15] for more details.

III. PROPOSED SOLUTIONS

The diagrams for two proposed solutions are shown in Figs. 2 and 3, where signals that are only used during DNN training are shown in green. Both methods employ a Feed-Forward (FF) DNN with two hidden layers and ReLU activation function. The number of nodes per layer is provided in Fig. 4.

A. TID Method

The non-linear signal, $\boldsymbol{\nu}_{r;n}$, transmitted by the relay is reconstructed using DNN (Fig. 2). The reconstructed non-linear signal, $\hat{\boldsymbol{\nu}}_{r;n}$, is used in the Least Squares (LS)

	Number of nodes per layer			
Method	inp. layer	1st hidden layer	2nd hidden layer	out layer
TID	2	6	4	2
TVD	8	10	4	2

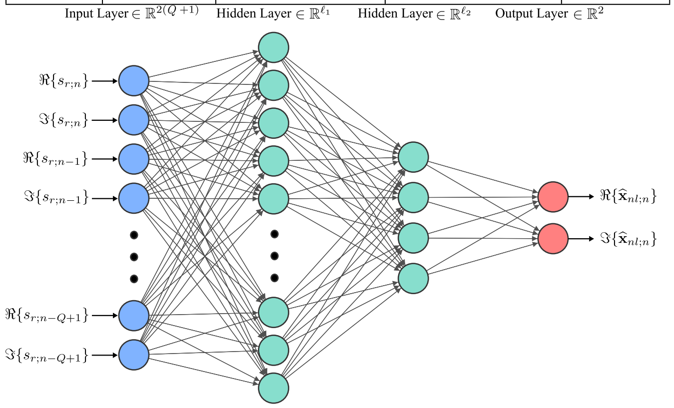

Fig. 4. A two-layer Feed-Forward Neural Network.

expression (6) for estimating the SI propagation channel as

$$\widehat{\mathbf{h}}^{(i)}_{rr;n} = \left(\widehat{\boldsymbol{\mathcal{V}}}^{H}_{r;\,n}\mathbf{C}^{(i)^{-1}}_{y_r}\widehat{\boldsymbol{\mathcal{V}}}_{r;n}\right)^{-1}\widehat{\boldsymbol{\mathcal{V}}}^{H}_{r;n}\mathbf{C}^{(i)^{-1}}_{y_r}\mathbf{y}^{(i)}_{r;\,n}, \quad (6)$$

where $\widehat{\boldsymbol{\mathcal{V}}}_{r;\,n}$ is a filtering matrix composed of signal samples reconstructed by the DNN, and $\mathbf{C}^{(i)}_{y_r}$ is the noise and interference covariance matrix defined as

$$\mathbf{C}^{(i)}_{y_r} = \varepsilon_s \mathbf{H}^{(i)}_{sr;\,n}\mathbf{H}^{(i)}_{sr;n}{}^{H} + \mathbf{I}\sigma^{(i)}_{w_r}{}^{2}. \quad (7)$$

Next, the SI component of the received signal is reconstructed using (5). The non-linear distortion introduced at the relay transmitter is hardware dependent, but assumed to be fixed for a given RF chip. Since reconstruction of the transmitted signal is solely based on the pre-distortion signal, $\mathbf{s}_{r;n}$, and does not depend on the propagation channel, $\mathbf{H}_{rr}$, the DNN can be trained off-line in advance, even if the SI propagation channel is time-varying. The training is performed using the input $\mathbf{s}_{r;n}$ and the non-linear output signal $\boldsymbol{\nu}_{\text{training}}$, which is available only in a factory setting.

B. TVD Method

A DNN is used for estimating the nonlinear components of the SI, while the linear SI is obtained using a Least Squares (LS) estimation technique, as shown in Fig. 3. We propose a Transfer Learning based two-stage DNN training strategy that significantly reduces on-line training overhead. The SI signal $\mathbf{x}^{(i)}_{rr;n}$ at the FD relay receiver in (5) can be decomposed as $\mathbf{x}^{(i)}_{rr;n} = \mathbf{x}^{(i)}_{lin;n} + \mathbf{x}^{(i)}_{nl;n}$, where $\mathbf{x}^{(i)}_{lin;n}$ is the linear SI signal, and $\mathbf{x}^{(i)}_{nl;n}$ contains the nonlinear and image SI components. First, we estimate the linear SI, $\widehat{\mathbf{x}}^{(i)}_{lin;n}$, while considering $\mathbf{x}^{(i)}_{nl;n}$ as part of the noise and interference. The linear SI cancellation signal $\widehat{\mathbf{x}}^{(i)}_{lin;n}$ is subtracted from the SI to obtain $\mathbf{x}^{(i)}_{nl;n} \approx \mathbf{x}^{(i)}_{rr;n} - \widehat{\mathbf{x}}^{(i)}_{lin;n}$, which is used for DNN training. The goal of the DNN is to reconstruct the nonlinear SI component $\widehat{\mathbf{x}}^{(i)}_{nl;n}$ using the subset of baseband samples $s_{r;n}$. The DNN has $2(M+1)$ input nodes, which correspond to the real and imaginary parts of the $(M+1)$ delayed version of the FD relay baseband sample $s_{r;n}$. The two output nodes represent the real and imaginary parts of the target nonlinear SI sample $\widehat{\mathbf{x}}^{(i)}_{nl;n}$. To overcome the DNN training limitations, we propose a two-phase training strategy based on the concept of *Transfer Learning*. First, we train our DNN model using training symbols generated at the FD relay node when there is no scheduled transmission between the source and destination nodes. We denote this stage as *pretraining*. During the actual source-destination transmission, we again perform training on the *pretrained* DNN model using a small number of training symbols. This second stage is called *finetuning*, which is performed to compensate the SI channel and nonlinear characteristic mismatch between *pretraining* and actual transmission time. Using this two-stage approach, the proposed SI canceller can achieve improved SI suppression with a small amount of training symbols and processing time. *Transfer Learning* is widely used in computer vision, [16].

IV. DNN Training Set Size Analysis

The following analysis focuses on the case where a DNN is used for modeling a time-invariant non-linear distortion. According to [17], the modeling error consists of two parts.

1) *Approximation Error:* a piece-wise linear approximation error, that approaches zero as the number of layers and nodes per layer increase (DNN-structure-dependent).
2) *Generalization Error:* is due to a poor coverage by the training set, i.e., a discrepancy between the statistics of the training set and those of the run-time data.

It was shown in [17] that ReLU DNN modeling of a nonlinear system reduces to a piece-wise linear approximation of a non-linear function. Hence, if the training data does not have at least k samples within some of the piece-wise linear segments, the weight and bias values determining these segments will not be properly selected ($k = 2$ since a straight line is fully defined by two points). If the run-time input samples happen to fall into these poorly-defined segments, the produced DNN output, $\hat{\nu}_r(m)$, in Fig. 2 will be incorrect, resulting in a generalization error. The following proposition defines an upper bound for the probability of such generalization error.

Proposition 1: If DNN is used for modeling a fixed nonlinear system, and both training and run-time DNN input samples, $s_r(m)$, have the same probability density function, $f(s_r(m))$, then an upper bound on the probability of generalization error, e_g, is given by

$$\mathbb{P}(e_g) \leq \sum_{l=1}^{N_{\text{seg,UB}}-1}\sum_{i=0}^{k-1}\binom{N_{tr}}{i}p_l^{i+1}(1-p_l)^{N_{tr}-i}, \quad (8)$$

where p_l is the probability of input sample $s_r(m)$ falling within the segment l (bounded by u_{l-1} and u_l).

In (8), k represents the minimum number of distinct training samples that is sufficient to properly define this linear segment ($k < N_{tr}$). According to [18], the maximal number of linear segments in a ReLU DNN is bounded by,

$$N_{\text{seg,LB}} \leq N_{\text{seg,max}} \leq N_{\text{seg,UB}}, \quad (9)$$

where

$$N_{\text{seg,UB}} = 2^{N_{\text{nodes}}}, \quad N_{\text{nodes}} = \sum_{j=1}^{\mathcal{L}} n_j, \quad (10)$$

$$N_{\text{seg,LB}} = \left(\prod_{j=1}^{\mathcal{L}-1}\left\lfloor\frac{n_j}{n_0}\right\rfloor^{n_0}\right)\sum_{m=0}^{n_0}\binom{n_{\mathcal{L}}}{m}, \quad (11)$$

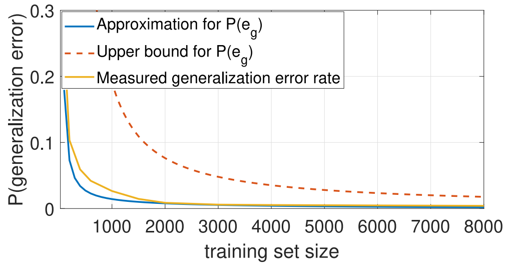

Fig. 5. Probability of generalization error as a function of training set size (2 hidden layers with 6 and 4 nodes, respectively).

where $\mathcal{L}$ is the number of hidden layers, n_j is the number of nodes in the j^{th} hidden layer, n_0 is the number of nodes in the input layer.

Proof: Assuming negligibly small probability of run-time sample being equal to the sole training sample in segment l, the probability of generalization error, e_g, is defined as,

$$\mathbb{P}(e_g) = \sum_{l=1}^{N_{seg}-1} \mathbb{P}(k-1; N_{tr}, p_l) \cdot p_{l,run}, \quad (12)$$

where the factor $p_{l,run} = p_l$ represents the probability of run-time sample falling into the segment l and is defined as,

$$p_l = \mathbb{P}(u_{l-1} \leq s_r(m) \leq u_l) = \int_{u_{l-1}}^{u_l} f(s_r(m))\, ds_r(m), \quad (13)$$

where $f(s_r(m))$ is PDF of $s_r(m)$. Given a training set of size N_{tr}, the probability of less than k training samples falling within segment l of $s_r(m)$ range is given by the following cumulative binomial distribution,

$$\mathbb{P}(k-1; N_{tr}, p_l) = \sum_{i=0}^{k-1} \binom{N_{tr}}{i} p_l^i (1-p_l)^{N_{tr}-i}. \quad (14)$$

Substituting (14) into (12), we obtain,

$$\mathbb{P}(e_g) = \sum_{l=1}^{N_{\text{seg}}-1} \sum_{i=0}^{k-1} \binom{N_{tr}}{i} p_l^i (1-p_l)^{N_{tr}-i} \cdot p_l. \quad (15)$$

Result (15) is not very useful since we do not have a way for finding N_{seg}. Since $N_{tr} > k$, the dominant factor in (15) is $(1-p_l)^{N_{tr}-i}$. Hence,

$$\mathbb{P}(e_g) \propto (1-p_l), \text{if } N_{tr} > k. \quad (16)$$

According to (13), the larger the segment size, $u_l - u_{l-1}$, the higher the value of p_l, i.e., $p_l \propto 1/N_{seg}$. Hence, the highest $\mathbb{P}(e_g)$ occurs when the number of segments is the largest, i.e., using (12), we arrive at

$$\mathbb{P}(e_g) \propto N_{seg} \leq N_{\text{seg,UB}}. \quad (17)$$

Replacing N_{seg} by $N_{\text{seg,UB}}$ in (15), results in (8). ∎

Using (11) in place of N_{seg} in (15), we obtain,

$$\mathbb{P}(e_g) \approx \sum_{l=1}^{N_{\text{seg,LB}}-1} \sum_{i=0}^{k-1} \binom{N_{tr}}{i} p_l^{i+1} (1-p_l)^{N_{tr}-i}. \quad (18)$$

As shown in Fig. 5, (18) provides a reasonable approximation of the true measured $\mathbb{P}(e_g)$. Equation (18) can be used as a guidance for selecting the training set size for a FF ReLU DNN (replacing trial and error approach). Fig. 5 is based on the TID method.

TABLE II
ON-LINE TRAINING COMPARISON

Method	Off-line training sample count	On-line training sample count
Polynomial Method [4]	0	30000
DNN-based Method [8]	0	20000
TVD Method	2760	276
TID Method	8000	0

V. SIMULATION RESULTS

A. Simulation Model Assumptions

We use the cubic polynomial non-linear distortion model from [14], where the following parameters are assumed: IIP3 = 13dB, linear gain = 10dB. The image rejection ratio in expression (4) is IRR $= 20\log_{10}(a_1/a_2) = 25$dB. Frequency-selective Rayleigh fading channel of order $M = 3$ is used. Same channel orders but different realizations were used for all propagation channels, i.e., $\mathbf{H}_{sr}$, $\mathbf{H}_{rd}$, and $\mathbf{H}_{rr}$. A new random channel realization was generated every data packet (packet size = 276 symbols). Blind channel estimation with fractionally-spaced MMSE equalization were employed at the relay and destination for $\mathbf{H}_{sr}$, $\mathbf{H}_{rd}$, respectively. The detailed treatment of this part can be found in [13]. A DNN with two hidden layers shown in Fig. 4 was employed in both methods. Different number of nodes for input and hidden layers were used for each solution. Both methods used the Adam minimization algorithm for training with mean-squared error (MSE) as the objective function. The DNN was implemented using the Keras framework with a TensorFlow backend.

B. Experimental Setup for the TID Method

Performance results discussed in Section V-D were obtained using only off-line training with 8000 training sets, where each set consists of two input samples, $\{\Re(s_{r,n}), \Im(s_{r,n})\}$, and the corresponding output samples, $\{\Re(\nu_{r,n}), \Im(\nu_{r,n})\}$. The training was performed over 15000 epochs.

C. Experimental Setup for the TVD Method

Neural Network: We considered a 2-layer FF DNN for the nonlinear cancellation (Fig. 4). The number of input layer nodes is $2(M+1) = 8$ (based on the SI channel order, M).

Pretraining: The learning rate of $\lambda = 0.001$, and a mini-batch size of $B = 12$. All remaining parameters have their default values. We generated an off-line data set of 10 distinct training packets. Using each of the training packets, we train the DNN model until the training loss converges.

Finetuning: We dedicate a single data packet for *finetuning* the DNN model trained in the *pretrained* stage. For *finetuning*, we use 90% symbols of the packet, while the remaining 10% is used for validation. The optimization algorithm and the parameters are identical to the *pretraining* stage.

D. Simulation Results and Discussion

As summarized in Table II, the proposed DNN-based SI estimation methods significantly reduce the number of on-line training symbols relative to Polynomial-based estimator [4] and DNN-based estimator [8]. This improvement is important for the overall system throughput since data transmission has to be stopped during the on-line training process. SI suppression performance of the proposed techniques (as a function of the number of on-line training epochs) is depicted in Fig. 6.

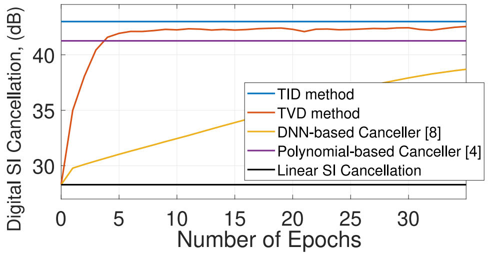

Fig. 6. Digital SI cancellation in dB w.r.t number of on-line epochs.

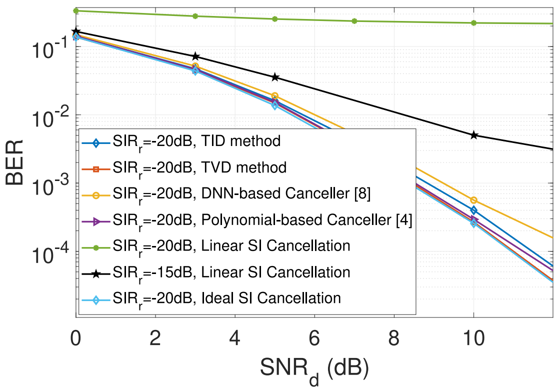

Fig. 7. BER plots for non-linear DNN-based and linear SI cancellation.

The polynomial-based SI cancellation of [4] and DNN-based nonlinear cancellation of [8] are also included for comparison. According to the results of Fig. 6, the proposed TID and TVD methods achieve 43dB and 42dB of SI suppression at 0 and 4 on-line training epochs, respectively. The proposed methods outperform [8], which requires 35 epochs to provide 39dB of SI reduction. They also outperform the nonlinear polynomial approach, which provides 41dB cancellation. Also, Fig. 6 shows that linear technique achieves only 28dB of SI cancellation. Therefore, the transfer-learning-based canceller is capable of superior SI cancellation with limited number of training symbols and training epochs. As shown in Fig. 7, BER performance of the proposed methods is close to that of an ideal digital SI cancellation. They provide lower BER than [8] and significantly outperform purely linear SI estimation.

VI. Conclusion

We presented two digital SI cancellation approaches for a FD relay link using DNN architecture for estimating the nonlinear SI signal. Using link-level simulation, it was demonstrated that both proposed SI mitigation techniques outperform the existing DNN-based methods. In the proposed TID method, the DNN-based reconstruction of non-linear distortion is decoupled from estimation of the linear SI propagation channel. This allowed complete elimination of the on-line training. The proposed TVD Method utilizes transfer learning and is capable of superior SI cancellation with limited number of on-line training symbols and training epochs. It can handle both time varying non-linear distortion as well as the SI propagation channel. An upper bound and an approximation for the probability of generalization error were derived. The approximation expression for the probability of generalization error (as a function of training set size), proved to be a close predictor of the measured error rate. Hence, the expression can be used for selecting the appropriate training set size.

Acknowledgment

The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the NSF.

References

[1] K. E. Kolodziej, B. T. Perry, and J. S. Herd, "In-band full-duplex technology: Techniques and systems survey," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 7, pp. 3025–3041, Jul. 2019.

[2] M. Heino *et al.*, "Recent advances in antenna design and interference cancellation algorithms for in-band full duplex relays," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 91–101, May 2015.

[3] S. Khaledian, F. Farzami, B. Smida, and D. Erricolo, "Inherent self-interference cancellation for in-band full-duplex single-antenna systems," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 6, pp. 2842–2850, Jun. 2018.

[4] D. Korpi, L. Anttila, and M. Valkama, "Nonlinear self-interference cancellation in MIMO full-duplex transceivers under crosstalk," *EURASIP J. Wireless Commun. Netw.*, Dec. 2017, Art. no. 24.

[5] M. A. Islam and B. Smida, "A comprehensive self-interference model for single-antenna full-duplex communication systems," in *Proc. IEEE ICC*, Shanghai, China, May 2019, pp. 1–7.

[6] L. Anttila, D. Korpi, E. Antonio-Rodriguez, R. Wichman, and M. Valkama, "Modeling and efficient cancellation of nonlinear self-interference in MIMO full-duplex transceivers," in *Proc. IEEE GLOBECOM*, Austin, TX, USA, Dec. 2014, pp. 777–783.

[7] Z. Zhang, Y. Shen, S. Shao, W. Pan, and Y. Tang, "Full duplex 2×2 MIMO radios," in *Proc. Int. Conf. Wireless Commun. Signal Process.*, Hefei, China, Oct. 2014, pp. 1–6.

[8] A. Balatsoukas-Stimming, "Non-linear digital self-interference cancellation for in-band full-duplex radios using neural networks," in *Proc. IEEE SPAWC*, Kalamata, Greece, Jun. 2018, pp. 1–5.

[9] H. Guo, S. Wu, H. Wang, and M. Daneshmand, "DSIC: Deep learning based self-interference cancellation for in-band full duplex wireless," in *Proc. IEEE GLOBECOM*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.

[10] C. Shi, Y. Hao, Y. Liu, and S. Shao, "Digital self-interference cancellation for full duplex wireless communication based on neural networks," in *Proc. Int. Conf. Commun. Inform. Syst.*, Wuhan, China, Dec. 2019, pp. 53–57.

[11] Y. Kurzo, A. Burg, and A. Balatsoukas-Stimming, "Design and implementation of a neural network aided self-interference cancellation scheme for full-duplex radios," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, Pacific Grove, CA, USA, Oct. 2018, pp. 589–593.

[12] K. Muranov, B. Smida, and N. Devroye, "On channel equalization for full-duplex relay networks," in *Proc. IEEE ICC*, Paris, France, May 2017, pp. 1–6.

[13] K. Muranov, B. Smida, and N. Devroye, "On blind channel estimation in full-duplex wireless relay systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4685–4701, Jul. 2021.

[14] Mathworks. *Memoryless Non-Linearity, Cubic Polynomial*. Accessed: Feb. 1, 2021. [Online]. Available: www.mathworks.com/help/comm/ref/memorylessnonlinearity.html

[15] J. Du, C. Jiang, J. Wang, Y. Ren, and M. Debbah, "Machine learning for 6G wireless networks: Carrying forward enhanced bandwidth, massive access, and ultrareliable/low-latency service," *IEEE Veh. Technol. Mag.*, vol. 15, no. 4, pp. 122–134, Dec. 2020.

[16] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*, vol. 1. Cambridge, MA, USA: MIT Press, 2016.

[17] H. Qiang, G. Feifei, Z. Hao, J. Shi, and L. G. Ye, "Deep learning for MIMO channel estimation: Interpretation, performance, and comparison," 2019. [Online]. Available: arXiv:1911.01918.

[18] G. F. Montufar, R. Pascanu, K. Cho, and Y. Bengio, "On the number of linear regions of deep neural networks," in *Advances in Neural Information Processing Systems*, vol. 27. Red Hook, NY, USA: Curran Assoc., 2014, pp. 2924–2932.