Machine Learning Techniques for Self-Interference Cancellation in Full-Duplex Systems

by

©Mohamed Elsayed Mohamed Selim

A dissertation submitted to the School of Graduate Studies
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Faculty of Engineering and Applied Science
Memorial University

May 2024

St. John's, Newfoundland

Abstract

Full-duplex (FD), enabling remote parties to transfer information simultaneously in both directions and in the same bandwidth, has been envisioned as an important technology for the next-generation wireless networks. This is due to the ability to leverage both time and frequency resources and theoretically double the spectral efficiency. Enabling the FD communications is, however, highly challenging due to the self-interference (SI), a leakage signal from the FD transmitter (Tx) to its own receiver (Rx). The power of the SI is significantly higher when compared with the signal of interest (SoI) from a remote node due to the proximity of the Tx to its co-located Rx. The SI signal is thus swamping the SoI and degrading the FD system's performance.

Traditional self-interference cancellation (SIC) approaches, spanning the propagation, analog, and/or digital domains, have been explored to cancel the SI in FD transceivers. Particularly, digital domain cancellation is typically performed using model-driven approaches, which have proven to be effective for SIC; however, they could impose additional cost, hardware, memory, and/or computational requirements.

Motivated by the aforementioned, this thesis aims to apply data-driven machine learning (ML)-assisted SIC approaches to cancel the SI in FD transceivers—in the digital domain—and address the extra requirements imposed by the traditional methods. Specifically, in Chapter 2, two grid-based neural network (NN) structures, referred to as ladder-wise grid structure and moving-window grid structure, are proposed to model the SI in FD transceivers with lower memory and computational requirements

than the literature benchmarks. Further reduction in the computational complexity is provided in Chapter 3, where two hybrid-layers NN structures, referred to as hybrid-convolutional recurrent NN and hybrid-convolutional recurrent dense NN, are proposed to model the FD SI. The proposed hybrid NN structures exhibit lower computational requirements than the grid-based structures and without degradation in the SIC performance. In Chapter 4, an output-feedback NN structure, referred to as the dual neurons-ℓ hidden layers NN, is designed to model the SI in FD transceivers with less memory and computational requirements than the grid-based and hybrid-layers NN structures and without any additional deterioration to the SIC performance.

In Chapter 5, support vector regressors (SVRs), variants of support vector machines, are proposed to cancel the SI in FD transceivers. A case study to assess the performance of SVR-based approaches compared to the classical and other ML-based approaches, using different performance metrics and two different test setups, is also provided in this chapter. The SVR-based SIC approaches are able to reduce the training time compared to the NN-based approaches, which are, contrarily, shown to be more efficient in terms of SIC, especially when high transmit power levels are utilized.

To further enhance the performance/complexity of the ML approaches provided in Chapter 5, two learning techniques are investigated in Chapters 6 and 7. Specifically, in Chapter 6, the concept of residual learning is exploited to develop an NN structure, referred to as residual real-valued time-delay NN, to model the FD SI with lower computational requirements than the benchmarks of Chapter 5. In Chapter 7, a fast and accurate learning algorithm, namely extreme learning machine, is proposed to suppress the SI in FD transceivers with a higher SIC performance and lower training overhead than the benchmarks of Chapter 5. Finally, in Chapter 8, the thesis conclusions are provided and the directions for future research are highlighted.

To my lovely father and mother ...
To my lovely wife and my coming kid, Hamza ...
To my entire family and friends ...
I hope this work makes you all proud ...

Acknowledgments

Praise be to Allah, the Cherisher and Sustainer of the Worlds. After thanking Almighty "ALLAH" for his blessing and guidance in completing this work, I would like to offer my sincere thanks to my supervisor, Prof. Octavia A. Dobre, for her invaluable guidance and support throughout my Ph.D. journey. I would like to express my deep appreciation to Prof. Dobre for her professionalism and personal caring for her students before and after the study; I was lucky to be her student, and I encourage the hard-working graduate students to go through this amazing experience with her. I would like to also thank Dr. Ahmad A. Aziz El-Banna for his invaluable guidance and contributions to complete this work. I would also like to acknowledge the financial support provided by my supervisor, the Faculty of Engineering and Applied Science, and the School of Graduate Studies.

I would like to thank Sohag University, my university in Egypt, for letting me try this fruitful experience at Memorial University. I would also like to thank my M.Sc. thesis's supervisors, Prof. Usama Sayed Mohammed and Prof. Hany S. Hussein, who taught me the foundations and put me on the right path of scientific research.

The final word of acknowledgment is reserved for my sweethearts, my parents, Mr. Elsayed and Mrs. Amal Selim, for their constant love and prayers, to my sisters, Shimaa, Dalia, and Menna, for their support, to my heartbeat and lovely wife, Yasmin, for her unbounded love, and to my coming kid, Hamza, "*I hope you see the light soon.*"

Co-Authorship Statement

I, Mohamed Elsayed Mohamed Selim, hold a principle author status for all the manuscript chapters (Chapters 2-7) in this dissertation. However, each manuscript is co-authored by my supervisor and co-researchers, whose contributions have facilitated the development of this work, as described below.

- Paper 1 in Chapter 2: M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Low complexity neural network structures for self-interference cancellation in full-duplex radio," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 181–185, Jan. 2021.

  I was the primary author, with authors 2-5 contributing to the idea, its formulation, and development, as well as the refinement of the presentation.

- Paper 2 in Chapter 3: M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Hybrid-layers neural network architectures for modeling the self-interference in full-duplex systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6291–6307, Jun. 2022.

  I was the primary author, with authors 2-5 contributing to the idea, its formulation, and development, as well as the refinement of the presentation.

- Paper 3 in Chapter 4: M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Full-duplex self-interference cancellation using dual-neurons neu-

ral networks," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 557–561, Mar. 2022.
I was the primary author, with authors 2-5 contributing to the idea, its formulation, and development, as well as the refinement of the presentation.

- Paper 4 in Chapter 5: M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Machine learning-based self-interference cancellation for full-duplex radio: Approaches, open challenges, and future research directions," *IEEE Open J. Veh. Technol.*, vol. 5, no. 1, pp. 21–47, Nov. 2023.
  I was the primary author, with authors 2-5 contributing to the idea, its formulation, and development, as well as the refinement of the presentation.

- Paper 5 in Chapter 6: M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Residual neural networks for learning the full-duplex self-interference," in *Proc. 57th Asilomar Conf. Signals, Syst.*, Comput., Nov. 2023, pp. 1–4.
  I was the primary author, with authors 2-5 contributing to the idea, its formulation, and development, as well as the refinement of the presentation.

- Paper 6 in Chapter 7: M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Extreme learning machine-assisted full-duplex self-interference cancellation," *to be submitted to IEEE Trans. Instrum. Meas.*, May 2024.
  I was the primary author, with authors 2-5 contributing to the idea, its formulation, and development, as well as the refinement of the presentation.

Mohamed Selim — May 21, 2024

Mohamed Selim — Date

Table of Contents

List of Tables

List of Figures

List of Abbreviations

1T1R	One transmit and one receive
2D	Two-dimensional
2HLNN	Two-hidden layers neural network
6G	Sixth-generation
ADC	Analog-to-digital converter
ADDs	Additions
APSM	Adaptive projected subgradient method
B5G	Beyond the fifth-generation
BPF	Band pass filter
BS	Batch size
CasCor NN	Cascade correlation neural network
CHRNN	Channel-robust neural network
CRELU	Complex rectified linear unit
CSID	Canonical system identification

CV-FFNN	Complex-valued feed-forward neural network
CV-TDNN	Complex-valued time delay neural network
DAC	Digital-to-analog converter
DL	Deep learning
DN-2HLNN	Dual-neurons two-hidden layers neural network
DN-LHLNN	Dual neurons-L hidden layers neural network
DR	Dynamic regression
DU	Deep unfolding
ELM	Extreme learning machine
EM	Expectation-maximization
FD	Full-duplex
FDD	Frequency division duplex
FFNN	Feed-forward neural network
FLOPs	Floating-point operations
FTRL	Follow the regularized leader
GMMs	Gaussian mixture models
HCRDNN	Hybrid convolutional recurrent dense neural network
HCRNN	Hybrid convolutional recurrent neural network
HD	Half-duplex

IBFD	In-band full-duplex
IMD2	Second-order intermodulation distortion
IQ	In-phase and quadrature-phase
LL	Lazy learning
LMS	Least mean-squares
LNA	Low-noise amplifier
LO	Local oscillator
LPF	Low pass filter
LR	Learning rate
LS	Least-squares
LWGS	Ladder-wise grid structure
MIMO	Multiple-input multiple-output
ML	Machine learning
MSE	Mean square error
MULs	Multiplications
MWGS	Moving-window grid structure
NN	Neural network
NTDSVR	Nested time-delay support vector regressor
OF	Output-feedback

OF-TDSVR	Output-feedback time-delay support vector regressor
OFDM	Orthogonal frequency division multiplexing
PA	Power amplifier
PC	Personal computer
PH	Parallel-Hammerstein
PSD	Power spectral density
QPSK	Quadrature phase-shift keying
RBF	Radial basis function
RELU	Rectified linear unit
Res-NN	Residual neural network
Res-RV-TDNN	Residual real-valued time-delay neural network
RF	Radio frequency
RFFs	Random Fourier features
RMSprop	Root mean square propagation
RNN	Recurrent neural network
RTDSVR	Residual time-delay support vector regressor
RV-FFNN	Real-valued feed-forward neural network
RV-TDNN	Real-valued time delay neural network
Rx	Receiver

SDR	Software-defined radio
SGD	Stochastic gradient descent
SI	Self-interference
SIC	Self-interference cancellation
SISO	Single-input single-output
SNR	Signal-to-noise ratio
SoI	Signal of interest
SVM	Support vector machine
SVR	Support vector regressor
TC	Tensor completion
TDNN	Time-delay neural network
TDSVR	Time-delay support vector regressor
TRP	Transmit receive point
TRX	Transceiver
Tx	Transmitter
VGA	Variable gain amplifier

Chapter 1

Introduction

1.1 Background

The sixth-generation (6G) wireless networks are anticipated to connect "intelligence" rather than "things" while maintaining the quality-of-service requirements of low latency, massive connectivity, and stringent energy efficiency [1]-[5]. Through several technologies, 6G visionaries expect an unprecedented provision of services to 6G users by allowing 10 times lower latency, 100 times higher connectivity, and 1000 times higher data rates compared to the fifth-generation wireless systems' users [2]-[4].

To meet the high data rate requirements of 6G networks, the in-band full-duplex (IBFD) systems have emerged as one of the potential technologies—as illustrated in Fig. 1.1 [6]—owing to their ability to serve a large number of devices concurrently on the same frequency bands [7]-[9]. Given this potential, the IBFD devices can theoretically provide a twofold increase in spectral efficiency, making them promising candidates for 6G networks. This is in contrast to half-duplex (HD) systems, in which only one-fold is provided by enabling a one-direction communication between remote parties, i.e., remote parties can either transmit or receive at the same time and in

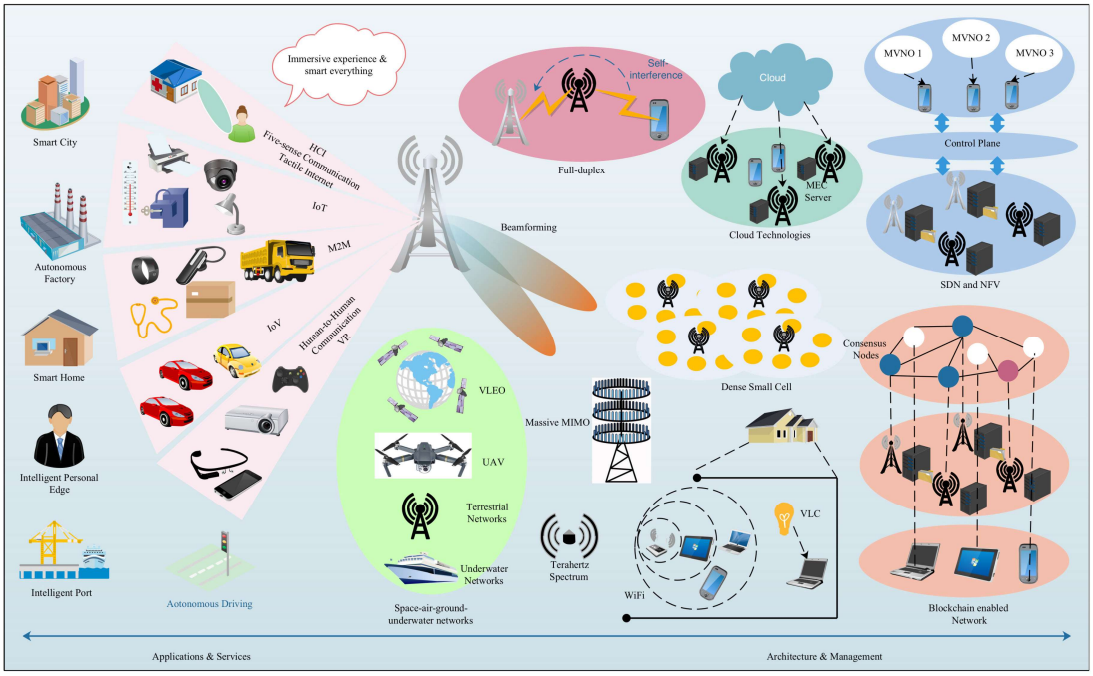


Fig. 1.1: Schematic diagram of future 6G wireless systems [6].

the same frequency band, as depicted in Fig. 1.2. Doubling the spectral efficiency offered by IBFD systems, however, comes at the cost of having an inevitable self-interference (SI) at the receiver (Rx) chain of an FD node from its own transmitter (Tx) chain, as illustrated in Fig 1.3. The power of the SI signal is significantly higher when compared with the desired signal of interest (SoI) from a remote node as a result of the proximity of the Tx to its co-located Rx. The SI is thus swamping the SoI and degrading the IBFD system's overall performance. To break through such a bottleneck, SI cancellation (SIC) has been verified as the panacea that can enable the essence of IBFD communications [7]-[9].

In the past few decades, researchers have drawn attention to canceling the SI in IBFD systems. Generally, the SIC can be performed in propagation, analog, and/or digital domains. Propagation domain cancellation can be performed at the radio frequency level using antenna isolation [7], beamforming [8], polarized antennas [10],

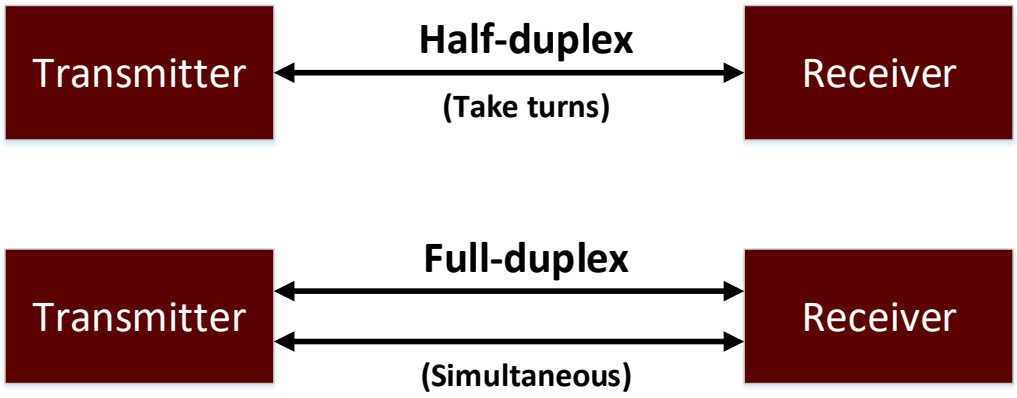


Fig. 1.2: HD versus FD communication.

circulators [11], and/or hybrid junction networks [12]. On the other hand, analog domain cancellation can be carried out actively by generating a pre-processed copy of the SI signal—using digitally-assisted auxiliary transmit chains, as an example—which is exploited to cancel the original SI signal at the Rx chain. Analog domain cancellations are often incapacitated to suppress the SI signal to the Rx noise floor level. As a consequence, additional focus has been directed to canceling the SI at the baseband level using digital domain cancellation [13]-[16]. At low or moderate transmit power levels, the digital domain cancellation is typically performed using linear cancelers, which construct an estimated copy of the SI signal based on techniques such as least-squares channel estimation [13], [14], [16]. However, at high transmit power levels, such linear cancellation becomes insufficient to entirely suppress the SI to the Rx noise floor due to the stringent non-linear behavior of the FD transceiver's components, such as the power and low-noise amplifiers [13]-[15]. Thus, non-linear digital SIC is applied with the linear cancellation to bring the SI to the Rx noise floor level. The non-linear digital SIC is conventionally performed using model-driven approaches, e.g., polynomial models, which are shown to fit well in practice; however, they need many trainable parameters that, in turn, translate to high computational

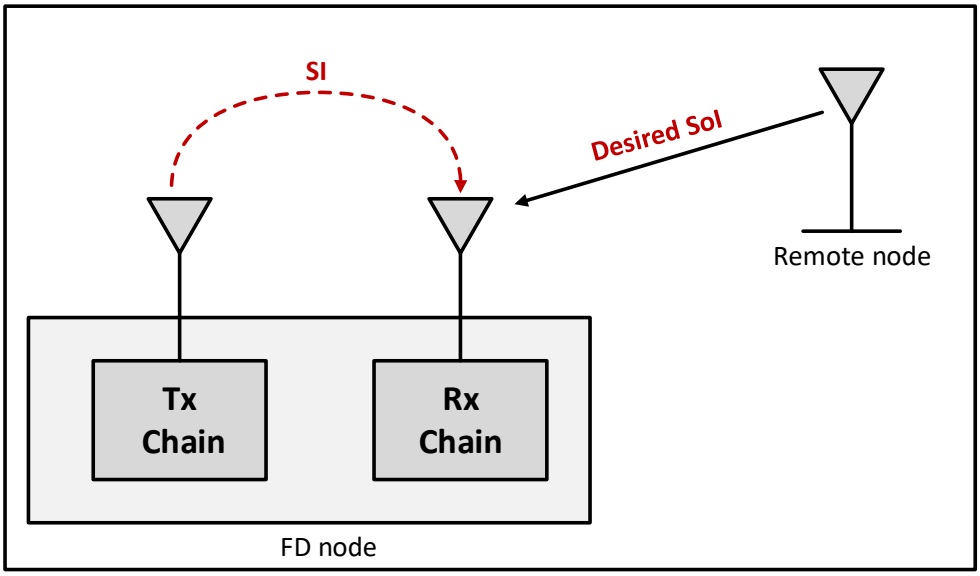


Fig. 1.3: SI in FD transceivers.

requirements [17].

Artificial intelligence, a wide-ranging area of computer science, has currently made a profound technological revolution in all disciplines of communications, as illustrated in Fig. 1.4 [18]-[30]. Specifically, machine and deep learning (ML and DL), aiming to extract hidden features, i.e., insights, from training data, have attained considerable success in channel coding [19], [20], channel estimation [22], [23], [26], [27], channel equalization [22], signal identification [21], [28], signal detection [18], optical fiber's signal-to-noise ratio (SNR) estimation [24], [25], digital pre-distortion [29], and power amplifier's behavioral modeling [30]. In these works, the data-driven ML approaches have achieved astonishing enhancements in performance and/or complexity when compared to the model-driven approaches.

Applying ML to IBFD communications has recently been regarded as one of the promising techniques that supports the horizon of 6G networks. To that extent, ML techniques, such as neural networks (NNs) [31]-[34], support vector regressors [35], [36], tensor completion [37], deep unfolding [38], and more [39]-[41], have been in-

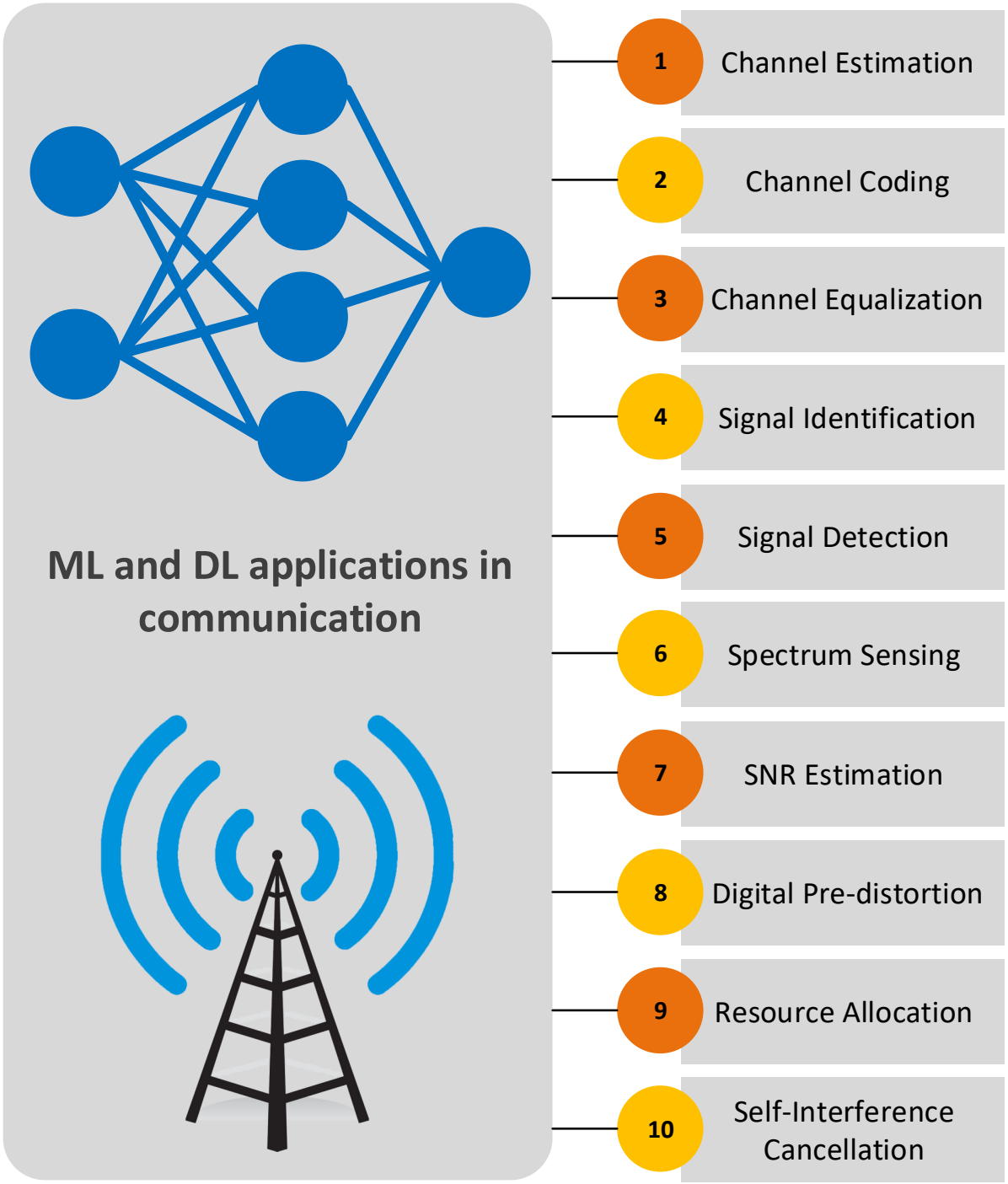


Fig. 1.4: ML and DL applications in communication.

troduced in the literature for canceling the SI and enabling the FD communications. Integrating ML with FD communications has achieved considerable success in terms of performance and/or complexity when compared to the traditional SIC approaches. However, further enhancements in performance and/or complexity are required to build energy-efficient ML-based cancelers, which can be suitable for hardware implementation in mobile communication platforms. Motivated by the above, this thesis aims to apply novel ML-assisted SIC approaches to cancel the SI in FD transceivers—in the digital domain—and address the extra requirements, e.g., memory and/or complexity, imposed by the traditional and existing ML-assisted SIC approaches.

In the next section, I will briefly review the traditional approaches for SIC in FD transceivers. Thereafter, the thesis motivation, contributions, and organization will be presented.

1.2 Traditional Approaches for SIC

Canceling the SI in FD transceivers can be performed using various techniques that span the propagation, analog, and/or digital domains [8], [9], as summarized in Fig. 1.5. The following subsections briefly review such SIC approaches, discussing their advantages, disadvantages, and/or challenges.

1.2.1 Propagation Domain Self-Interference Cancellation

Canceling the SI within the propagation domain is typically performed at the early stage of the FD transceiver, i.e., it revolves around the Tx and Rx antennas. Propagation domain cancellation can be accomplished passively using techniques such as antenna separation, coupling networks, phase control, cross-polarization, and/or surface treatments [8], [9], as shown in Fig. 1.5. Alternatively, it can be done actively

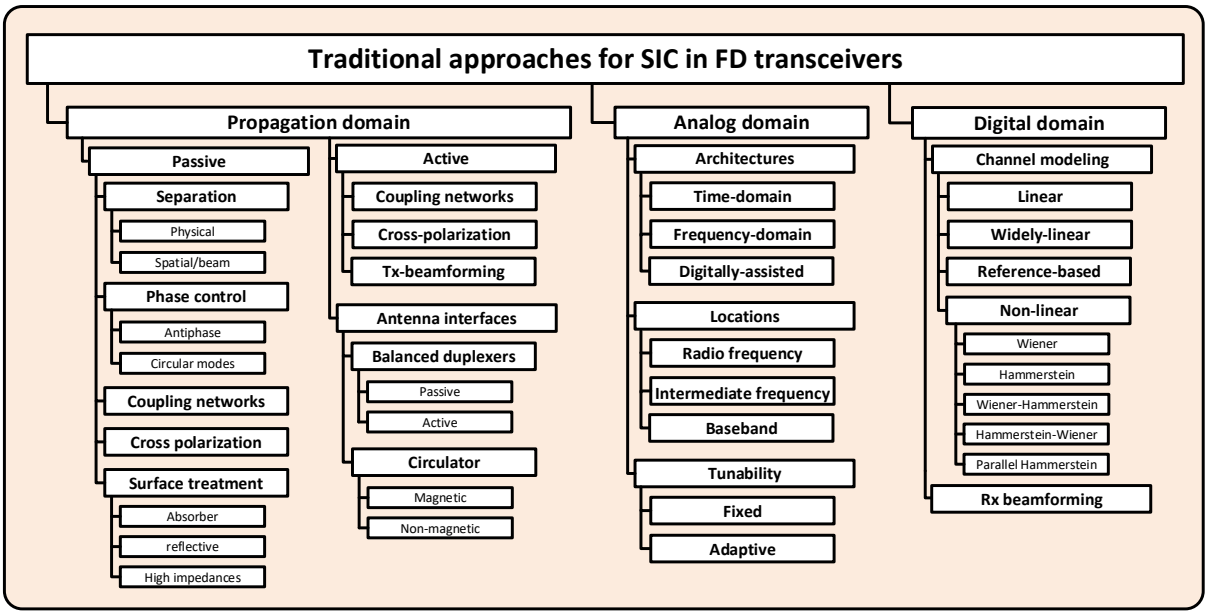


Fig. 1.5: Traditional approaches for SIC in FD transceivers [9].

using techniques such as active coupling networks, active cross-polarization, and/or Tx beamforming [9]. Additionally, antenna interfaces, such as balanced duplexers and circulators, can also be employed. Applying the SIC within the propagation domain has the advantage of refraining the SI signal from saturating the front end of the FD Rx; however, in some cases, it may lead to the suppression of the desired signal, i.e., the SoI [8]. Also, it can come at the cost of adding a hardware circuity to the FD transceiver. Hence, the focus is directed to additionally canceling the SI in other signal domains, e.g., analog and digital domains.

1.2.2 Analog Domain Self-Interference Cancellation

Canceling the SI within the analog domain is performed in the analog circuits between the antennas and digital conversion stages [8], [9]. Analog domain cancellation approaches can be classified based on their architecture, location, and tunability, as illustrated in Fig. 1.5 [9]. One of the common architectures for analog domain cancellation is to use digitally-assisted techniques based on auxiliary transmit chains [9]. Digitally assisted analog domain cancellation has the advantage of preventing the SI

signal from saturating the analog-to-digital converter (ADC), especially in mobility channel environments. However, the processing in the analog domain can be very costly and challenging to scale up into a higher number of antennas, i.e., multiple-input multiple-output scenario [8]. The focus is thus directed to additionally canceling the SI in the digital domain.

1.2.3 Digital Domain Self-Interference Cancellation

Canceling the SI in the digital domain is performed after the ADC using techniques such as channel modeling and/or Rx beamforming, as shown in Fig. 1.5. Digital domain approaches, applying channel modeling techniques, use the fact that the Rx of any IBFD node has knowledge of its transmitted signal in order to model the transceiver's impairments. Specifically, in channel modeling-based SIC, linear, widely linear, and reference-based models are applied to approximate the SI channel effects. Additionally, non-linear models, such as Wiener, Hammerstein, Wiener-Hammerstein, and parallel Hammerstein—the cornerstone of the widely utilized polynomial models—are employed to model the transceiver's non-linearities, as shown in Fig. 1.5. Digital domain cancellation has the advantage that the processing becomes relatively easy to perform and less hardware-expensive compared to the analog domain cancellation [8]; however, it can come at the cost of increasing the computational complexity of the FD transceiver [17].[1]

From the previous discussion, applying the traditional approaches for SIC in FD transceivers can come with challenges, such as imposing extra hardware, higher cost, and/or additional computational complexity. In contrast, applying ML approaches for SIC in FD communications can relax such requirements, as mentioned before.

[1]The traditional approaches for SIC are detailed in [8] and [9].

1.3 Thesis Motivation

As stated before, a considerable enhancement in terms of performance and/or complexity has been achieved by integrating ML with FD communications. However, further improvements in performance and/or complexity are needed to build energy-efficient ML-based cancelers, which can be suitable for hardware implementation in communication platforms. Motivated by the aforementioned, this thesis aims to apply novel ML-assisted SIC approaches to cancel the SI in FD transceivers—in the digital domain—and address the extra requirements imposed by the traditional and existing ML-assisted SIC approaches.

1.4 Thesis Contribution

Motivated by the aforementioned, in this thesis, I have identified and investigated the following research points:

1. I have designed two grid-based NN structures, referred to as ladder-wise grid structure (LWGS) and moving-window grid structure (MWGS), to model the SI in FD transceivers [42]. The optimum hyperparameters for the two NNs have been obtained. The computational complexity and memory storage requirements for the two algorithms have also been evaluated in terms of the number of floating-point operations (FLOPs) and network parameters, respectively, and analyzed compared to those of the literature benchmarks. The proposed LWGS and MWGS-based cancelers are able to model the FD SI with lower memory and computational requirements than the literature benchmarks.

2. I have proposed two hybrid-layers NN architectures, referred to as hybrid-convolutional recurrent NN (HCRNN) and hybrid-convolutional recurrent dense

NN (HCRDNN), to model the SI in FD transceivers [43]. The optimum settings for training the HCRNN and HCRDNN, e.g., number of convolutional filters, filter size, number of neurons in recurrent and dense layers, activation functions, learning rate, batch size, and optimizer, have been obtained. The computational and memory requirements of the proposed NNs have been derived. Performance analysis of the proposed NNs has also been provided in terms of their prediction capabilities, mean square error, achieved SIC, computational complexity, and memory requirements. The proposed NNs are able to model the SI with lower computational requirements than the existing and grid-based NN structures.

3. I have designed an output-feedback NN structure, referred to as dual neurons-ℓ hidden layers NN (DN-ℓHLNN), to model the SI in FD transceivers [44]. The network settings of the proposed algorithm have been optimized. The computational complexity and memory storage have been studied. The proposed DN-ℓHLNN is able to model the SI with lower memory and computational requirements than the existing, grid-based, and hybrid-layers NN architectures.

4. I have investigated the support vector regressors (SVRs), variants of support vector machines, to cancel the SI in FD transceivers [45]. In this study, a general and comprehensive system model to integrate ML with FD communications has been firstly introduced. The traditional and other ML approaches investigated for SIC have been reviewed. A case study to assess the performance of SVR-based approaches compared to the traditional and other ML-based approaches, using different performance metrics and two different test setups, has also been provided. An efficiency measure to select a suitable ML approach for SIC in FD transceivers, depending on the system demands, has also been devised in this chapter. The SVR-based SIC approaches are able to reduce the training time

compared to the NN-based approaches, which are, contrarily, shown to be more efficient in terms of SIC, especially when high transmit power levels are utilized.

5. I have exploited the concept of residual learning to develop an NN structure, referred to as residual real-valued time-delay NN (Res-RV-TDNN), to model the FD SI [46]. The optimum hyperparameters of the proposed NN have been obtained. Further, the computational complexity and memory storage of the proposed algorithm have been evaluated in terms of the number of FLOPs and network parameters, respectively. The proposed Res-RV-TDNN is able to model the SI with lower computational requirements than the benchmarks in [45].

6. I have proposed the extreme learning machine (ELM), a fast and accurate learning algorithm, to suppress the SI in FD transceivers with high SIC and low training overhead [47]. The network settings of the proposed ELM-assisted SIC approach have been optimized, and its performance has been analyzed compared to the benchmarks in [45] in terms of SIC, PSD, training overhead, memory storage, and computational complexity. The proposed ELM-assisted SIC approach is able to model the SI with higher SIC performance and lower training overhead than the benchmarks, albeit at higher memory and computational requirements.

1.5 Thesis Organization

The rest of this dissertation is organized as follows. Chapter 2 proposes grid-based NN structures for canceling the SI in FD transceivers. Chapter 3 proposes hybrid-layers NN architectures for suppressing the FD SI. Chapter 4 proposes an output-feedback NN structure for SIC in FD transceivers. Chapter 5 investigates the performance of SVR-based SIC approaches to cancel the SI in FD transceivers. Chapter 6 explores the concept of residual learning for SIC. Chapter 7 proposes an ELM-assisted SIC

approach to suppress the SI in FD transceivers. Finally, Chapter 8 draws the thesis's conclusions and provides a guide for future research directions.

References

[1] L. Bariah, L. Mohjazi, S. Muhaidat, P. C. Sofotasios, G. K. Kurt, H. Yanikomeroglu, and O. A. Dobre, "A prospective look: Key enabling technologies, applications and open research topics in 6G networks," *IEEE Access*, vol. 8, pp. 174792–174820, Aug. 2020.

[2] P. Yang, Y. Xiao, M. Xiao, and S. Li, "6G wireless communications: Vision and potential techniques," *IEEE Netw.*, vol. 33, no. 4, pp. 70–75, Jul. 2019.

[3] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?," *Nature Electron.*, vol. 3, no. 1, pp. 20–29, Jan. 2020.

[4] X. You *et al.*, "Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Sci. China Inf. Sci.*, vol. 64, no. 1, pp. 1–74, Nov. 2020.

[5] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, D. Niyato, O. A. Dobre, H. V. Poor, "6G internet of things: A comprehensive survey," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 359–383, Jan. 2022.

[6] F. Guo, F. R. Yu, H. Zhang, X. Li, H. Ji, and V. C. M. Leung, "Enabling massive IoT toward 6G: A comprehensive survey," *IEEE Internet Things J.*, vol. 8, no. 15, pp. 11891–11915, Aug. 2021.

[7] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

[8] 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.

[9] 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.

[10] B. Debaillie *et al.*, "Analog/RF solutions enabling compact full-duplex radios," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1662–1673, Sep. 2014.

[11] Y.-S. Choi and H. Shirani-Mehr, "Simultaneous transmission and reception: Algorithm, design and system level performance," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 12, pp. 5992–6010, Dec. 2013.

[12] L. Laughlin, M. A. Beach, K. A. Morris, and J. L. Haine, "Optimum single antenna full duplex using hybrid junctions," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1653–1661, Sep. 2014.

[13] D. Korpi, L. Anttila, V. Syrjala, and M. Valkama, "Widely linear digital self-interference cancellation in direct-conversion fullduplex transceiver," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1674–1687, Sep. 2014.

[14] D. Korpi, *et al*, "Digital self-interference cancellation under nonideal RF components: Advanced algorithms and measured performance," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2015, pp. 286–290.

[15] M. A. Tafreshi, M. Koskela, D. Korpi, P. Jääskeläinen, M. Valkama, and J. Takala, "Software defined radio implementation of adaptive nonlinear digital self-interference cancellation for mobile inband fullduplex radio," in *Proc. IEEE Global Conf. Signal Inform. Process. (GLOBALSIP)*, Dec. 2016, pp. 733–737.

[16] M. S. Amjad and O. Gurbuz, "Linear digital cancellation with reduced computational complexity for full-duplex radios," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[17] D. Korpi, L. Anttila, and M. Valkama, "Nonlinear self-interference cancellation in MIMO full-duplex transceivers under crosstalk," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, pp. 1–15, Dec. 2017.

[18] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[19] E. Nachmani, E. Marciano, L. Lugosch, W. J. Gross, D. Burshtein, and Y. Be'ery, "Deep learning methods for improved decoding of linear codes," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 119–131, Feb. 2018.

[20] F. Liang, C. Shen, and F. Wu, "An iterative BP-CNN architecture for channel decoding," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 144–159, Feb. 2018.

[21] M. Gao, Y. Li, O. A. Dobre, and N. Al-Dhahir, "Joint blind identification of the number of transmit antennas and MIMO schemes using Gerschgorin radii and FNN," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 373–387, Jan. 2019.

[22] X. Cheng, D. Liu, C. Wang, S. Yan, and Z. Zhu, "Deep learning-based channel estimation and equalization scheme for fbmc/oqam systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 881–884, Jun. 2019.

[23] M. Xu, S. Zhang, J. Ma, and O. A. Dobre, "Deep learning-based time-varying channel estimation for RIS assisted communication," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 94–98, Jan. 2022.

[24] M. Al-Nahhal, I. Al-Nahhal, O. A. Dobre, X. Lin, D. Chang, and C. Li, "Joint estimation of linear and nonlinear coherent optical fiber signal to-noise ratio," *IEEE Photon. Technol. Lett.*, vol. 35, no. 1, pp. 23–26, Nov. 2022.

[25] M. Al-Nahhal, I. Al-Nahhal, O. A. Dobre, S. K. O. Soman, D. Chang, and C. Li, "Learned signal-to-noise ratio estimation in optical fiber communication links," *IEEE Photon. J.*, vol. 14, no. 6, pp. 1–7, Nov. 2022.

[26] Y. Liu, I. Al-Nahhal, O. A. Dobre, and F. Wang, "Deep-learning-based channel estimation for IRS-assisted ISAC system," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2022, pp. 4220–4225.

[27] Y. Liu, I. Al-Nahhal, O. A. Dobre, and F. Wang, "Deep-learning channel estimation for IRS-assisted integrated sensing and communication system," *IEEE Trans. Veh Technol.*, early access, pp. 1–14, Dec. 2022.

[28] E. A. Makled, I. Al-Nahhal, O. A. Dobre, and O. Üreten, "Identification of cellular signal measurements using machine learning," *IEEE Trans. Instrum. Meas.*, vol. 72, no. 1, pp. 1–4, Jan. 2023.

[29] 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.

[30] 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.

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

[32] 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. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 589–593.

[33] 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.

[34] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Advanced machine learning techniques for self-interference cancellation in full-duplex radios," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 1149–1153.

[35] M. Erdem, H. Ozkan, and O. Gurbuz, "Nonlinear digital self-interference cancellation with SVR for full duplex communication," in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC)*, Jun. 2020, pp. 1–6.

[36] M. Yilan, O. Gurbuz, and H. Ozkan, "Integrated linear and nonlinear digital cancellation for full duplex communication," *IEEE Wireless Communications*, vol. 28, no. 1, pp. 20–27, Feb. 2021.

[37] F. Jochems and A. Balatsoukas-Stimming, "Non-linear self-interference cancellation via tensor completion," in *Proc. 54th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2020, pp. 905–909.

[38] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Identification of nonlinear RF systems using backpropagation," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2020, pp. 1–6.

[39] M. Erdem, H. Ozkan, and O. Gurbuz, "A new online nonlinear self-interference cancelation method with random Fourier features," *IEEE Wireless Commun. Lett.*, vol. 11, no. 7, pp. 1379–1383, Apr. 2022.

[40] J. Chen, L. Zhang, and Y.-C. Liang, "Exploiting Gaussian mixture model clustering for full-duplex transceiver design," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5802–5816, Aug. 2019.

[41] O. Zhao, W-S Liao, K. Li, T. Matsumura, F. Kojima, and H. Harada, "Lazy learning-based self-interference cancellation for wireless communication systems with in-band full-duplex operations," in *Proc. IEEE 32nd Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2021, pp. 1589–1594.

[42] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Low complexity neural network structures for self-interference cancellation in full-duplex radio," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 181–185, Jan. 2021.

[43] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Hybrid-layers neural network architectures for modeling the self-interference in full-duplex systems," *IEEE Trans. Veh Technol.*, vol. 71, no. 6, pp. 6291–6307, Jun. 2022.

[44] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Full-duplex self-interference cancellation using dual-neurons neural networks," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 557–561, Mar. 2022.

[45] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Machine learning-based self-interference cancellation for full-duplex radio: Approaches, open challenges, and future research directions," *IEEE Open J. Veh. Technol.*, vol. 5, no. 1, pp. 21–47, Nov. 2023.

[46] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Residual neural networks for learning the full-duplex self-interference," in *Proc. 57th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2023, pp. 1–4.

[47] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Extreme learning machine-assisted full-duplex self-interference cancellation," *to be submitted IEEE Trans. Instrum. Meas.*, May 2024.

Chapter 2

Low Complexity Neural Network Structures for Self-Interference Cancellation in Full-Duplex Radio

2.1 Abstract

Self-interference (SI) is considered as a main challenge in full-duplex (FD) systems. Therefore, efficient SI cancelers are required for the influential deployment of FD systems in beyond fifth-generation wireless networks. Existing methods for SI cancellation have mostly considered the polynomial representation of the SI signal at the receiver. These methods are shown to operate well in practice while requiring high computational complexity. Alternatively, neural networks (NNs) are envisioned as promising candidates for modeling the SI signal with reduced computational complexity. Consequently, in this chapter, two novel low complexity NN structures, referred to as the ladder-wise grid structure (LWGS) and moving-window grid structure (MWGS), are proposed. The core idea of these two structures is to mimic the non-

linearity and memory effect introduced to the SI signal in order to achieve proper SI cancellation while exhibiting low computational complexity. The simulation results reveal that the LWGS and MWGS NN-based cancelers attain the same cancellation performance of the polynomial-based canceler while providing 49.87% and 34.19% complexity reduction, respectively.

2.2 Introduction

The recent advancements in wireless technology impose a tremendous increase in the number of devices that are required to satisfy the ascending demand for high data rates communication. This high increase leads to an undeniable fact that some levels of saturation in the available frequency resources will be reached. Therefore, using efficient methods for sharing the spectrum resources is eagerly mandated for the next generations of wireless systems, such as beyond the fifth-generation [1]. Full-duplex (FD) technology has emerged as a promising remedy for spectrum congestion problem by providing an efficient way for spectrum sharing. In FD systems, the data is transmitted and received at the same time slot and in the same band of frequency [2]. Sharing the spectrum resources simultaneously has the potential of doubling the spectral efficiency of FD systems. However, this in turn, gives rise to a substantial problem known as the self-interference (SI), which occurs when the transmitter's signal is leaked into the FD receiver. As such, canceling the SI signal at the receiver is deemed the main challenge against the practical deployment of FD systems [3], [4].

For typical FD systems, the SI signal could be 110 dB larger than the desired signal of interest at the receiver [5]. Therefore, if not efficiently eliminated, the SI signal may saturate the receiver's analog components, such as the analog-to-digital converter (ADC) and the low-noise amplifier (LNA) [2]. Existing methods for SI can-

cellation employ analog domain cancellation techniques, which are performed either passively using the physical separation between the transmit and receive antennas or actively by injecting a cancellation waveform into the propagation path of the received signal [3]. However, the analog cancellation techniques are not usually able to completely eliminate the SI signal at the receiver. Hence, the residual amount of the SI signal is further suppressed with the help of digital domain cancellation [6]. For that, an estimated SI signal is subtracted from the received signal to perform the SI cancellation. The digital cancellation procedure seems to be an easy task in theory; however, it is hard to be realized in practice due to the non-linear distortion caused by the various parts of the transceiver, such as the power amplifier (PA), IQ mixer, ADC, and digital-to-analog converter (DAC) [7]. This distortion makes the SI signal entirely different from the digital transmitted signal and raises a challenge for the perfect elimination of the SI signal at the receiver. Typically, the polynomial model is used for modeling the non-linearities caused by different parts of the FD transceiver. The polynomial model works properly in practice while suffering from high computational complexity [8].

Recently, neural networks (NNs) have received remarkable research interest from communication community experts due to their advantages in modeling the non-linearities with reduced computational complexity [8]–[11]. In [8], the authors introduce a real-valued feed-forward NN (RV-FFNN) to model the SI signal. Further, the hardware implementation of this NN-based canceler is presented in [10]. The same research group proposes the complex-valued FFNN (CV-FFNN) to perform the SI cancellation, and shows that the CV-FFNN could achieve the same cancellation as the RV-FFNN with a reduced number of floating-point operations (FLOPs) [11]. In addition, in [11], the recurrent NN (RNN) is introduced for SI cancellation due to its capability to model data sequences; it has been shown that the RNN is not a proper

candidate solution for the SI cancellation problem due to its high computational complexity. To the best of our knowledge, the research works of [8] and [11] represent the few attempts that target the application of NNs for SI cancellation in FD systems, and there is a scarcity of contributions in this field.

Subsequently, in this chapter, two novel NN structures, referred to as the ladder-wise grid structure (LWGS) and moving-window grid structure (MWGS), are proposed. The aim of these structures is to model the SI signal with low computational complexity. The proposed NNs exploit a grid topology in which only partial connections among the different neurons in the input and hidden layers are utilized to model the SI signal with reduced computational complexity. In addition, the proposed methods aim to learn the memory effect introduced to the SI signal in order to efficiently perform the SI cancellation. The numerical simulations substantiate the validity of the proposed LWGS and MWGS NN-based cancelers as they achieve the same cancellation performance of the polynomial canceler while providing 49.87% and 34.19% reduction in the number of FLOPs, respectively. Besides, the proposed LWGS and MWGS outperform the state-of-the-art NN-based cancelers in terms of computational complexity.

2.3 Full-Duplex System Model

The system model of the FD transceiver is depicted in Fig. 2.1. In this chapter, the polynomial model is used to approximate the SI signal. Therefore, I follow the stipulated assumption in [7] that the IQ mixer and PA are considered the dominant sources of non-linearities in the FD transceiver. Accordingly, the non-linear effect of other transceiver components, such as the DAC, ADC, variable gain amplifier (VGA), and LNA, is neglected. Furthermore, due to the use of a shared local oscillator (LO)

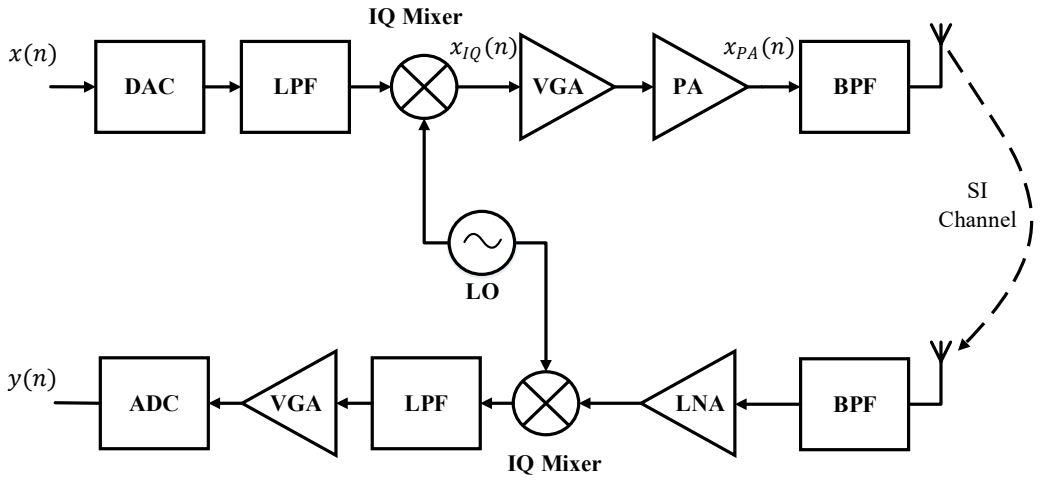


Fig. 2.1: Full-duplex transceiver system model.

for both transmitter and receiver, the effect of the phase noise is also ignored [10]. As such, the digital transmitted signal $x(n)$ is firstly converted from the digital to analog form using the DAC. The analog signal is then filtered using a low pass filter (LPF) and mixed with the carrier signal at the IQ mixer. The IQ mixer adds a non-linear distortion to the input signal due to the IQ imbalance, and the digital equivalent of the IQ mixer signal can be written as [7]

$$x_{IQ}(n) = \frac{1}{2}(1 + \psi e^{j\theta})\, x(n) + \frac{1}{2}(1 - \psi e^{j\theta})\, x^*(n), \tag{2.1}$$

where ψ and θ denote the transmitter's gain and phase imbalance parameters, respectively. The mixed signal is then amplified using the PA, which further distorts the input signal by adding additional non-linearities. In this chapter, I consider the parallel Hammerstein model to approximate the non-linear distortion of the PA [7]. Subsequently, the output signal of the PA can be expressed as follows [7], [8]:

$$x_{PA}(n) = \sum_{\substack{p=1,\\ p\,odd}}^{P} \sum_{m=0}^{M_{PA}} h_{m,p}\, x_{IQ}(n-m)^{\frac{p+1}{2}} x_{IQ}^*(n-m)^{\frac{p-1}{2}}, \tag{2.2}$$

where $h_{m,p}$ represents the impulse response of the parallel Hammerstein model, while P and M_{PA} are the non-linearity order and memory length of the PA, respectively.

The amplified signal is then leaked into the receiver via the SI channel forming the SI signal. With the assumption that the FD system does not receive any signal from any remote FD nodes (i.e., no signal of interest is considered) and there is no thermal noise, only the SI signal will go through the receiver. The received SI signal is firstly filtered by the band pass filter (BPF), then amplified by the LNA, down-converted by means of the IQ mixer, and finally converted to digital form using the ADC. The SI signal at the receiver output is expressed as

$$y(n) = \sum_{\substack{p=1,\\ p\,odd}}^{P} \sum_{q=0}^{p} \sum_{m=0}^{M-1} h_{m,q,p}\, x(n-m)^q x^*(n-m)^{p-q}, \tag{2.3}$$

where $h_{m,q,p}$ indicates the impulse response of a channel including the overall effect of the PA, IQ mixer, and SI channel, while M denotes the memory effect introduced to the input signal by the PA and SI channel.

The main aim of the digital canceler is to generate an accurate estimated version $\hat{y}(n)$ of the SI signal $y(n)$ at the receiver. Therefore, to perform the digital SI cancellation, $\hat{y}(n)$ is subtracted from $y(n)$, and the residual amount of the SI is approximated as $y_r(n) = y(n) - \hat{y}(n)$. Hence, the amount of SI cancellation can be given in dB as

$$\Psi_{dB} = 10 \log_{10} \left(\sum_n |y(n)|^2 \,/\, \sum_n |y_r(n)|^2 \right). \tag{2.4}$$

2.4 Proposed NN-Based Cancelers

Cascade forward NN is an NN architecture that utilizes additional connections from the input and every pre-layer to every post-layer [12]. The cascade forward NN has

been broadly utilized to model the time series data, and it is shown to work well in a wide variety of problems [12]. A similar NN that employs a cascade structure is the cascade correlation NN (CasCor NN) [13], a promising solution to speed up the learning algorithms of the conventional NNs, such as the back-propagation. CasCor NN starts with a simple network topology that contains only input and output units, and then successively adds hidden units one by one until the desired level of network error is accomplished. The resulting network is formed in a grid topology in which each new added unit is connected to the input and other layers' units in a cascade structure fashion.

CasCor NN has a faster learning capability than the conventional NNs that apply back-propagation algorithms. Moreover, it is not mandatory in CasCor NN to determine the network structure before the training phase since the network automatically determines its optimum configuration [13]. However, the major disadvantage of CasCor NN is that it potentially overfits to the training data in the sense that it yields a better performance on the training set while achieving worse performance on a previously unseen (i.e., new) data [14]. As a result, modified versions of CasCor NN have been introduced to avoid the overfitting of CasCor NN by applying simplified grid structures with only partial connections in the grid [15]. Based on this, for the SI cancellation, various grid structures can be investigated to model the memory effect introduced to the SI signal in order to achieve a desired cancellation performance with reduced computational complexity.

Motivated by this promising idea, two novel low complexity NN structures, named as the LWGS and MWGS, are proposed. The proposed NNs employ a grid topology with partial connections among the different neurons in the input and hidden layers. The main difference between the LWGS and MWGS lies in the way utilized by each structure to pass the buffered samples of the input signal to the different neurons

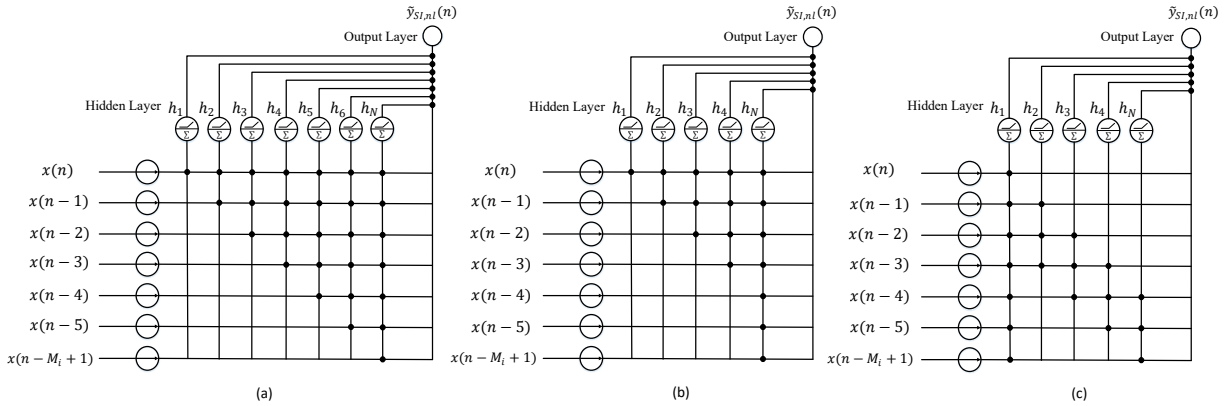


Fig. 2.2: Proposed NNs (a) LWGS for $N = M$ (b) LWGS for $N < M$ (c) MWGS.

in the grid in order to efficiently simulate the SI signal's memory effect. The key ideas and network structures of the proposed methods are presented in detail in the following subsections.

2.4.1 Ladder-Wise Grid Structure (LWGS)

To imitate the memory effect introduced to the SI signal, the LWGS is proposed as shown in Figs. 2.2(a) and (b). The LWGS employs a grid structure similar to that used in the CasCor NN. However, the LWGS uses the standard back-propagation technique to minimize the network's error and cannot determine its own structure as the CasCor NN. Accordingly, in the LWGS, the network structure is selected empirically before training to achieve the target network performance.

The basic idea behind the LWGS is to feed the buffered data to the network neurons in a stair-case manner as depicted in Figs. 2.2(a), (b). Here, I denote the number of hidden units by N. As such, in Fig. 2.2(a), I consider the case when the number of hidden units is equal to the number of input units (e.g., $N = M = 7$), where M is the memory length as stated before. Starting with the stair base, the instantaneous sample $x(n)$ is passed to all the neurons, and every predecessor sample

(i.e., $x(n-1)$, $x(n-2)$, ... etc.) is passed to a fewer number of neurons gradually. The oldest sample, $x(n-M+1)$, which is the least one related to the current sample $x(n)$, is allowed to be passed to only one neuron side by side with its followers. In this manner, each neuron receives the instantaneous sample plus part of the buffered samples to learn the temporal behavior of the SI signal, and the outputs of all neurons are then combined to figure out the detected pattern. Following this approach, the LWGS could model the SI signal with only partial connections in the grid, and therefore it can result in a significant reduction in the computational complexity.

Furthermore, the LWGS can learn the memory effect introduced to the SI signal with fewer connections between the input and hidden layers' neurons. Reducing the number of connections can be done by considering a shorter length of the ladder base (i.e., reducing the number of neurons to be less than the memory length ($N < M$)) as shown in Fig. 2.2(b). The idea of this configuration is to enable the recent delayed samples that are more related to the instantaneous sample $x(n)$ to be learned using many neurons; however, the other samples that are less related to $x(n)$ are learned by only one neuron (e.g., $x(n-4)$, $x(n-5)$, $x(n-M+1)$) in Fig. 2.2(b). This will slightly degrade the performance of the LWGS while providing a significant reduction in the computational complexity compared to the case when $N = M$.

In the proposed method, the SI cancellation is performed in the digital domain in which the non-linear part of the digital SI cancellation signal is reconstructed using the LWGS canceler. However, the linear part of the cancellation signal is estimated using the conventional least square channel estimation technique where all the non-linear effects of the different transceiver's components are neglected [8]. The total cancellation achieved by the LWGS canceler is then computed by summing the linear and non-linear cancellations [8], [10].

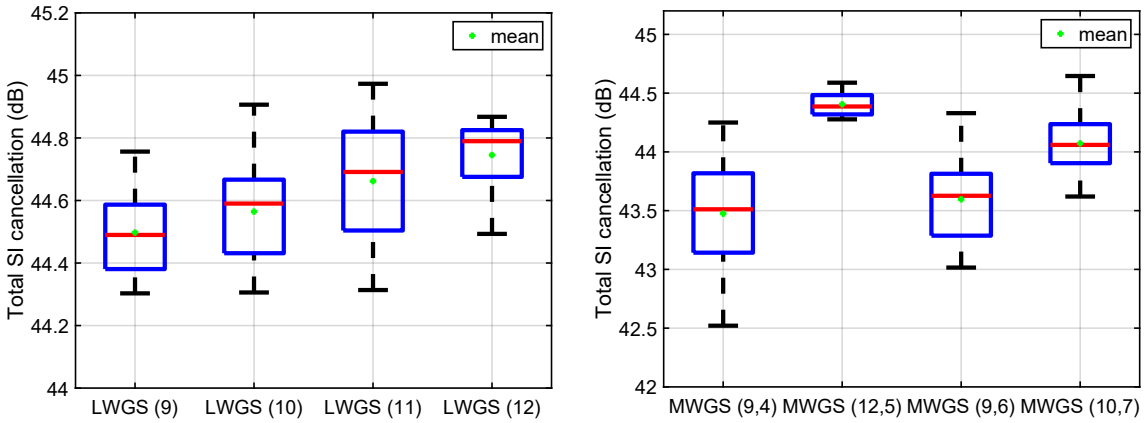
The effect of varying the number of hidden layer's neurons on the cancellation

performance of the LWGS is shown in Fig. 2.3(a).[1] I test the LWGS using $N = 9, 10, 11, 12$ and depict the boxplots of total cancellation achieved by various configurations using 20 seed initializations. The LWGS shows flexible settings that suit different applications. For example, moving from LWGS with nine neurons (i.e., LWGS (9)) to twelve neurons (i.e., LWGS (12)) augments the SI cancellation from 44.50 to 44.75 dB; however, the increased number of neurons would result in an increased computational complexity.

2.4.2 Moving-Window Grid Structure (MWGS)

An alternative approach that can accommodate the memory effect of the SI signal is the moving window technique, generally recognized as an effective method for time series prediction [16]. As such, I consider the moving window with a grid topology to form the MWGS. Similar to the LWGS, the MWGS applies the standard back-propagation technique to minimize the network's error. Further, the MWGS takes advantage of the reduced connections in the network grid. However, in the MWGS, the considered samples of the input signals learned by different neurons are partitioned based on a fixed-length sliding window technique as depicted in Fig. 2.2(c). More specifically, all the input samples are passed to the first neuron. The main purpose of this neuron is to learn the dependencies between all the delayed samples of the input signal. Moreover, the other employed neurons are allowed to assist in learning the memory effect by considering the windowed data only. Besides, sliding the window over different neurons allows to consider all the buffered samples caused by the non-linearities of the aforementioned FD components. For example, in Fig. 2.2(c), a window size $W = 3$ is employed. Therefore, the first, second, and third delayed version of $x(n)$ (i.e., $x(n-1)$, $x(n-2)$, $x(n-3)$) are considered by the second neuron,

[1]The results in Fig. 2.3 are obtained using the simulation parameters in Section 2.6.



(a) LWGS versus different number of neurons.

(b) MWGS versus different number of neurons and window sizes.

Fig. 2.3: SI cancellation boxplots.

while $x(n-2)$, $x(n-3)$, and $x(n-4)$ samples are recognized by the third neuron, and so forth.

The effect of varying the number of hidden layer's neurons N and the window size W on the cancellation performance of the MWGS is studied as well.[1] In this study, the network structures of the MWGS are selected empirically based on a trial and error approach. As such, I test the values of $N = 9, 10, 11, 12$ and $W = 4, 5, 6, 7$. Due to having many combinations between N and W, in Fig. 2.3(b), I only show the best four structures that achieve the highest SI cancellation. As seen from the figure, the MWGS using twelve neurons and window size $W = 5$ (i.e., MWGS (12,5)) attains the highest SI cancellation among the competing structures.

2.5 Computational Complexity

In this chapter, I consider the total number of FLOPs as an indicator of the computational complexity of the NN-based cancelers. The total number of FLOPs is evaluated by calculating the total number of real-valued operations used in the NN's

inference process as $\Xi = \Xi_{w,b} + \Xi_a$ where $\Xi_{w,b} = \xi_{w,b}^{(R_m)} + \xi_{w,b}^{(R_a)}$ is the sum of real-valued multiplications and additions, which account for multiplying the input/previous layer values by the weight matrix and adding the bias terms. Similarly, $\Xi_a = \xi_a^{(R_m)} + \xi_a^{(R_a)}$ represents the sum of real-valued multiplications and additions required to evaluate the activation functions in the hidden layer's neurons.

In the LWGS and MWGS NNs, the inputs, hidden layer values, and network parameters (e.g., weights and biases) are complex-valued numbers. Therefore, converting the complex-valued multiplications and additions to their real-valued equivalents is required. By employing the reduced multiplications approach [10], a complex-valued multiplication requires three real multiplications and five real additions. Moreover, since each complex-valued addition is implemented using two real additions, $\xi_{w,b}^{(R_m)}$ and $\xi_{w,b}^{(R_a)}$ can be expressed as

$$\xi_{w,b}^{(R_m)} = 3\xi_{w,b}^{(C_m)}, \tag{2.5}$$

$$\xi_{w,b}^{(R_a)} = 5\xi_{w,b}^{(C_m)} + 2\xi_{w,b}^{(C_a)}, \tag{2.6}$$

where $\xi_{w,b}^{(C_m)}$ and $\xi_{w,b}^{(C_a)}$ represent the number of complex-valued multiplications and additions, respectively, which account for handling the weights and biases operations.

In the LWGS, $\xi_{w,b}^{(C_m)}$ and $\xi_{w,b}^{(C_a)}$ can be calculated as

$$\xi_{w,b}^{(C_m)} = \xi_{w,b}^{(C_a)} = \sum_{i=1}^{N} i + M. \tag{2.7}$$

Further, in the MWGS, $\xi_{w,b}^{(C_m)}$ and $\xi_{w,b}^{(C_a)}$ can be obtained as

$$\xi_{w,b}^{(C_m)} = \xi_{w,b}^{(C_a)} = M + W(N-1) + N. \tag{2.8}$$

However, $\xi_{w,b}^{(C_m)}$ and $\xi_{w,b}^{(C_a)}$ for CV-FFNN can be expressed as

$$\xi_{w,b}^{(C_m)} = \xi_{w,b}^{(C_a)} = N(M+1). \tag{2.9}$$

The proposed LWGS and MWGS employ the complex rectified linear unit (CRELU) activation function, which is defined as [11]

$$\Phi(z) = \max(0, \Re(z)) + j \max(0, \Im(z)), \tag{2.10}$$

where $\Re(z)$ and $\Im(z)$ denote the real and imaginary parts of z, respectively. The implementation of CRELU activation function (2.10) requires two real multiplications and two complex additions (i.e., four real additions) to evaluate the real and imaginary parts of z. Further, to evaluate the $\max(0, \Re(z))$ and $\max(0, \Im(z))$, two multiplexers and two comparators are required. Herein, if I assume that each comparator comes with no cost and each multiplexer costs one real addition [8], the implementation of CRELU activation function requires two real multiplications and six real additions.[2] As such, the number of real-valued multiplications $\xi_a^{(R_m)}$ and additions $\xi_a^{(R_a)}$ utilized for evaluating the activation functions in the hidden layer's neurons of the LWGS and MWGS can be given by $\xi_a^{(R_m)} = 2N$ and $\xi_a^{(R_a)} = 6N$, respectively.

2.6 Results and Discussion

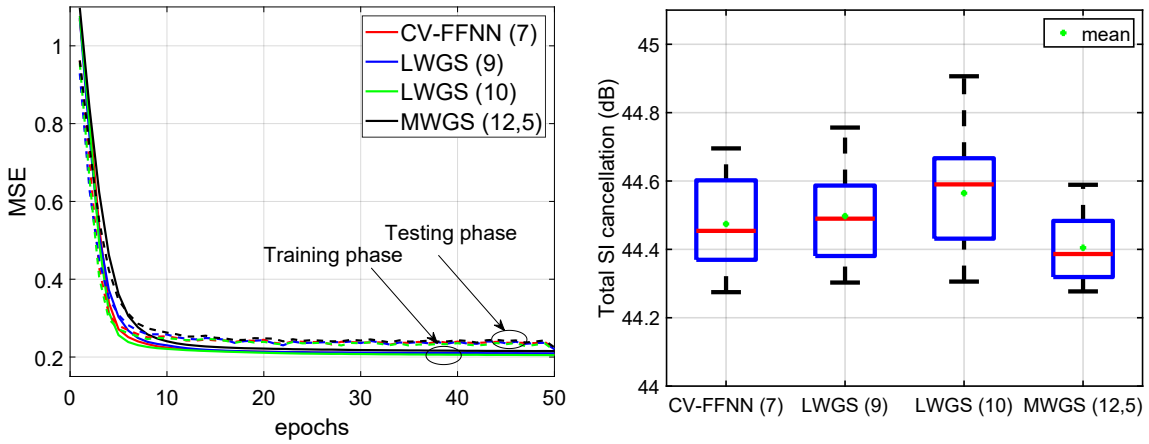
In this section, I assess the performance of the LWGS and MWGS NN-based cancelers in terms of mean square error (MSE), SI cancellation, and computational complexity. In addition, a comparison between the proposed methods and the NN-based cancelers in the literature is also investigated. All the considered NNs are trained using complex-

[2]It is noted that the activation functions' complexity in [11] is evaluated by counting their usage in the hidden layer's neurons, which is not exact.

valued inputs and implemented in Python using Keras library and TensorFlow back-end. In this work, I examine the use of the measured dataset presented in [8] and [11]. Hence, I train and test the NNs using measured data from a realistic FD testbed, which applies an orthogonal frequency division multiplexing signal with 1024 sub-carriers using a quadrature phase-shift keying modulation and 10 MHz pass-band bandwidth. The dataset, containing 20,480 samples, is split into a training set that consists of 90% of samples and a testing set that includes the remaining 10%. I adopt the back-propagation technique, Adam optimization, and CRELU activation function for the NNs [11]. The networks' hyperparameters, such as the batch size and learning rate, are tuned to select their optimal values. Based on hyperparameters tuning, I employ a learning rate of 0.0045 and a batch size of 62 to train the NNs. Besides, I consider $M = 13$ for the polynomial and NN-based cancelers [8], [11].

All the NN-based cancelers are employed to model the non-linear part of the SI signal. Furthermore, for the sake of comparison, the NNs settings are selected in such a way that they achieve a similar cancellation performance to the polynomial canceler with $P = 5$. The polynomial canceler at $P = 5$ produces 44.45 dB cancellation and requires 1556 FLOPs and 312 network parameters to be implemented [11]. As such, to achieve the target cancellation of the polynomial canceler, the CV-FFNN requires at least a single hidden layer with seven neurons (i.e., CV-FFNN (7)) [11]. In addition, from Fig. 2.3, it is observed that the LWGS (9) and LWGS (10) achieve the target cancellation as they provide 44.50 and 44.56 dB, respectively. Further, the MWGS (12,5) attains 44.40 dB, which is very close to the target cancellation. Thus, in this analysis, I consider CV-FFNN (7), LWGS (9), LWGS (10), and MWGS (12,5) as promising NN-based cancelers that can be used as alternatives to the traditional polynomial canceler.

In Fig. 2.4(a), the MSE values of the aforementioned NNs are evaluated on the



(a) MSE performance.

(b) SI cancellation boxplots.

Fig. 2.4: Performance comparison for different network structures.

training and testing data, respectively, using 20 seed initializations. As seen from the figure, the considered NNs achieve a comparable MSE for the target cancellation performance. Fig. 2.4(b) depicts the boxplots of SI cancellation achieved by the considered NN-based cancelers using the above-selected settings. It is apparent from the figure that CV-FFNN (7), LWGS (9), and MWGS (12,5) attain a comparable cancellation performance to the polynomial canceler. However, LWGS (10) provides a slightly higher SI cancellation. It is worth noting that the LWGS structure slightly outperforms the cancellation of the MWGS as it passes the instantaneous sample $x(n)$ (i.e., most significant sample) to all neurons, which enables it to learn the SI signal's temporal behavior better than the MWGS.

The complexity analysis for the different NN-based cancelers is provided in Table 2.1, where the polynomial canceler complexity is considered as reference, and the NNs complexity is computed in terms of the number of FLOPs and the number of network parameters used to perform the total SI cancellation (i.e., linear and non-linear cancellations). As seen from the table, the LWGS (9) reduces the number of FLOPs by more than 49% while achieving a similar cancellation to the polynomial-based

TABLE 2.1: Complexity reduction for different network structures compared to the polynomial model with $P = 5$.

Network	Cancellation	Complexity		Complexity Reduction	
		# Parameters	# FLOPs	# Parameters	# FLOPs
Polynomial ($P=5$)	44.45 dB	312	1556	-	-
CV-FFNN (7)	44.47 dB	238	1164	-23.72%	-25.19%
LWGS (9)	44.50 dB	162	780	-48.08%	-49.87%
LWGS (10)	44.56 dB	184	888	-41.03%	-42.93%
MWGS (12,5)	44.40 dB	212	1024	-32.05%	-34.19%

canceler. Furthermore, the LWGS (10) outperforms the cancellation performance of the polynomial canceler while requiring 7% more FLOPs than the LWGS (9). Accordingly, the proposed LWGS provides a flexible trade-off between the cancellation performance and the computational complexity. In addition, the MWGS (12,5) saves 34% computations compared to the polynomial-based canceler, while the conventional CV-FFNN (7) saves only 25% of the computations. The previous results reveal the superiority of the proposed NNs compared to the polynomial and state-of-the-art NN-based cancelers.

2.7 Conclusion

In this chapter, two novel low complexity NN structures, namely the ladder-wise grid structure (LWGS) and moving-window grid structure (MWGS), are proposed to model the SI signal with low computational complexity. The former employs a stair-based structure to accommodate the memory effect of the SI signal. The latter uses a fixed-window procedure to model the temporal behavior of the SI signal. Our findings showed that the proposed LWGS and MWGS provide the same cancellation performance of the polynomial-based canceler while attaining 49.87% and 34.19% reduction in the computational complexity, respectively. In addition, the proposed

LWGS and MWGS offer superior performance over the state-of-the-art NN-based cancelers by exhibiting 24.7% and 9% complexity reduction, respectively.

References

[1] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, Mar. 2020.

[2] 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.

[3] H. V. Nguyen et al., "On the spectral and energy efficiencies of fullduplex cell-free massive MIMO," *IEEE J. Sel. Areas Commun.,* vol. 38, no. 8, pp. 1698–1718, Aug. 2020.

[4] F. Chen, R. Morawski, and T. Le-Ngoc, "Self-interference channel characterization for wideband 2 × 2 MIMO full-duplex transceivers using dual-polarized antennas," *IEEE Trans. Antennas Propag.*, vol. 66, no. 4, pp. 1967–1976, Apr. 2018.

[5] G. Agrawal, S. Aniruddhan, and R. K. Ganti, "A compact mixer-first receiver with >24 dB self-interference cancellation for fullduplex radios," *IEEE Microw. Wireless Compon. Lett.*, vol. 26, no. 12, pp. 1005–1007, Dec. 2016.

[6] Z. Zhang, K. Long, A. V. Vasilakos, and L. Hanzo, "Full-duplex wireless communications: Challenges, solutions, and future research directions," *Proc. IEEE*, vol. 104, no. 7, pp. 1369–1409, Jul. 2016.

[7] D. Korpi, L. Anttila, and M. Valkama, "Nonlinear self-interference cancellation in MIMO full-duplex transceivers under crosstalk," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, pp. 1–15, Dec. 2017.

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

[9] Y. Yang, S. Zhang, F. Gao, J. Ma, and O. A. Dobre, "Graph neural network based channel tracking for massive MIMO networks," Apr. 2020. [Online]. Available: https://arxiv.org/abs/2004.08738v1.

[10] 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.

[11] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Advanced machine learning techniques for self-interference cancellation in full-duplex radios," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 1149–1153.

[12] B. Warsito, R. Santoso, and H. Yasin, "Cascade forward neural network for time series prediction," *J. Phys. Conf. Ser.*, vol. 1025, no. 1, pp. 1-9, May 2018.

[13] S. E. Fahlman and C. Lebiere, "The cascade-correlation learning architecture," in *Proc. Adv. Neural Inf. Process. Syst.*, Jun. 1990, pp. 524–532.

[14] F. J. Śmieja, "Neural network constructive algorithms: Trading generalization for learning efficiency?," *Circuits Syst., Signal Processing*, vol. 12, no. 2, pp. 331–374, Jun. 1993.

[15] T. Y. Kwok and D. Y. Yeung, "Constructive algorithms for structure learning in feedforward neural networks for regression problems," *IEEE Trans. Neural Networks*, vol. 8, no. 3, pp. 630–645, May 1997.

[16] D. Ruta, B. Gabrys, and C. Lemke, "A generic multilevel architecture for time series prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 3, pp. 350–359, Mar. 2011.

Chapter 3

Hybrid-Layers Neural Network Architectures for Modeling the Self-Interference in Full-Duplex Systems

3.1 Abstract

Full-duplex (FD) systems have been introduced to provide high data rates for beyond fifth-generation wireless networks through simultaneous transmission of information over the same frequency resources. However, the operation of FD systems is practically limited by the self-interference (SI), and efficient SI cancelers are sought to make the FD systems realizable. Typically, polynomial-based cancelers are employed to mitigate the SI; nevertheless, they suffer from high complexity. This chapter proposes two novel hybrid-layers neural network (NN) architectures to cancel the SI with low complexity. The first architecture is referred to as hybrid-convolutional recurrent NN

(HCRNN), whereas the second is termed as hybrid-convolutional recurrent dense NN (HCRDNN). In contrast to the state-of-the-art NNs that employ dense or recurrent layers for SI modeling, the proposed NNs exploit, in a novel manner, a combination of different hidden layers (e.g., convolutional, recurrent, and/or dense) in order to model the SI with lower computational complexity than the polynomial and the state-of-the-art NN-based cancelers. The key idea behind using hybrid layers is to build an NN model, which makes use of the characteristics of the different layers employed in its architecture. More specifically, in the HCRNN, a convolutional layer is employed to extract the input data features using a reduced network scale. Moreover, a recurrent layer is then applied to assist in learning the temporal behavior of the input signal from the localized feature map of the convolutional layer. In the HCRDNN, an additional dense layer is exploited to add another degree of freedom for adapting the NN settings in order to achieve the best compromise between the cancellation performance and computational complexity. The complexity analysis of the proposed NN architectures is provided, and the optimum settings for their training are selected. The simulation results demonstrate that the proposed HCRNN and HCRDNN-based cancelers attain the same cancellation of the polynomial and the state-of-the-art NN-based cancelers with an astounding computational complexity reduction. Furthermore, the proposed cancelers show high design flexibility for hardware implementation, depending on the system demands.

3.2 Introduction

Recently, the evolution of Internet-of-Everything, supporting massive connectivity among billions of users and billions of devices, has imposed a radical shift towards the next generation of wireless networks, such as beyond the fifth-generation (B5G)

[1]. These modern networks aim to provide high reliability, low latency, and high data rates, in the order of tens Gbits/s, to enable extended reality applications, live multimedia streaming, and autonomous systems in smart cities and factories, such as drone swarms, cars, and robotics [2], [3]. As such, to cater to this novel breed of applications and support such a plethora of services, the next generation of wireless systems should be inherently tailored to simultaneously deliver higher data rates with lower communication delays for both uplink and downlink.

In this regard, full-duplex (FD) has emerged as one of the key enabling technologies for B5G wireless networks by providing high data rates through simultaneous transmission of information over the same frequency resources [4]-[7]. Efficient exploitation of the resources enables the FD systems to meet the high quality-of-service requirements in terms of spectral efficiency, which represents a major factor in designing B5G wireless networks. Despite of this potential, the main challenge in implementing the FD systems is the self-interference (SI), which comes out from the transmitter of the same device on its own receiver [8]. This undesirable interference significantly hinders the proliferation of FD systems in the next generation of wireless networks [4]-[7].

Over the past decade, a flurry of research interest has been directed for canceling the interference in FD systems to make them realizable [7], [8]. Typically, canceling the SI can be implemented in analog radio frequency (RF) and/or digital domains to bring the SI signal's power down to the receiver's noise level. The analog RF suppression is implemented at the very first stage of the receiver chain to refrain the SI signal from saturating the analog components of the receiver, such as the low-noise amplifier (LNA), variable-gain amplifier (VGA), and analog-to-digital converter (ADC) [7]. In particular, the analog RF cancellation can be classified into passive and active cancellations [9], [10]. Passive cancellation is implemented using techniques such as antenna separation [11], circulators [12], polarized antennas [13], and balanced

hybrid-junction networks [14]. On the other side, the active suppression is performed using analog circuits, which generate a copy of the SI signal in order to be subtracted from the original SI signal at the receiver chain [10]. In general, the analog suppression techniques are insufficient to entirely remove the SI at the receiver side, and a non-negligible residual SI still exists after the analog cancellation process. Hence, digital cancellation approaches are utilized in order to mitigate the residual interference [15].

Digital domain cancellation uses the same notion of active suppression where a processed copy of the baseband transmitted signal is subtracted from the residual SI signal, but in the digital domain [15]. In principle, the digital cancelers could effectively eliminate this SI signal since it stems from a transmit signal that is obviously known to the receiver. However, this is not the case in practice, as the SI signal is significantly distorted by the SI coupling channel and the impairments of the transceiver components, such as non-linear distortion of the power amplifier (PA), in-phase and quadrature-phase (IQ) imbalance of the mixer, phase noise of imperfect transceiver's oscillators, and digital-to-analog converter (DAC) and ADC's quantization noise [16].

In order to efficiently cancel the SI in the digital domain, the digital cancelers should properly model the distortion incorporated into the input signal due to the imperfection of the hardware and the SI channel. Generally, modeling the transceiver's impairments is based on the polynomial approximation of the SI signal at the receiver side [15]. The polynomial-based models have excellent modeling capabilities to mimic the SI signal; however, they suffer from high complexity [16]. Accordingly, low-complexity modeling approaches are sought for approximating the SI signal in FD systems.

Applying neural networks (NNs) and deep learning has gained significant momentum in the field of signal processing and wireless communications in the last few years [17]-[28]. NNs have been recently employed to replace the model-based approaches in

numerous communication areas in order to approximate the non-linearities with good performance and low implementation complexity. For instance, NNs have brought breakthroughs in signal detection [17], signal classification [18], channel estimation [19], [20], channel equalization [21], channel coding [22], PA modeling [23]-[25], digital pre-distortion [24]-[27], and non-linearity compensation in optical fiber systems [28].

In addition, there has been a surge of interest in applying NNs for SI cancellation in FD systems [29]-[33]. More specifically, the first attempt of using NNs for canceling the SI has been reported in [29], where a real-valued time delay NN (RV-TDNN)[1] is introduced to model the SI signal with computational complexity lower than the polynomial-based canceler. In [30], a recurrent NN (RNN) and a complex-valued TDNN (CV-TDNN) have been investigated for SI mitigation; it is shown that the CV-TDNN has excellent modeling capabilities to approximate the SI with lower computational complexity than the polynomial and RNN-based cancelers. In [31], [32], the hardware design of the polynomial and NN-based cancelers introduced in [29] have been provided. Furthermore, in [33], the ladder-wise grid structure (LWGS) and moving-window grid structures (MWGS), two low-complexity NN models, have been introduced for SI cancellation. It is demonstrated that the LWGS and MWGS attain a similar cancellation performance to the polynomial and CV-TDNN-based cancelers with a significant complexity reduction. The previous works shed light on the few attempts that target applying low-complexity NN models for SI cancellation in FD systems. However, further enhancements in the complexity are required to build energy-efficient NN-based cancelers, which can be suitable for hardware implementation in mobile communication platforms. As such, this study fills in this gap by providing efficient NN-based SI cancelers, which achieve a similar cancellation perfor-

[1]It is noted that the RV feed-forward NN (RV-FFNN) in [29] has a similar structure to the RV-TDNN in [23] since both employ the input signal's buffered samples at the input layer. Henceforth, I will use RV-TDNN instead of RV-FFNN for accurate referring.

mance to that of the polynomial and the state-of-the-art NN-based cancelers while attaining a remarkable complexity reduction.

Based on the aforementioned, in this chapter, two novel low-complexity NN architectures referred to as the hybrid-convolutional recurrent NN (HCRNN) and hybrid-convolutional recurrent dense NN (HCRDNN) are proposed. The proposed NNs exploit hybrid hidden layers (e.g., convolutional, recurrent, and/or dense) to efficiently model the memory effect and non-linearity incorporated into the SI signal, with low complexity. The key idea behind using hybrid layers is to build an NN model, which makes use of the characteristics of the different layers employed in its architecture. In particular, the proposed NNs exploit, in a novel manner, the feature extraction characteristics of the convolutional layer along with the sequence modeling capabilities of the recurrent layer and/or the learning abilities of the dense layer in order to model the SI with lower computational complexity than the polynomial and the state-of-the-art NN-based cancelers. To the best of the author's knowledge, applying hybrid-layers NN architectures for SI cancellation has not been previously reported in the literature, and it is introduced for the first time in this chapter. More specifically, in the proposed HCRNN, the input data containing the I/Q components of the input samples is formulated into a two-dimensional (2D) graph for the sake of suitable processing by the convolutional layer. The convolutional layer is then applied to the 2D graph to extract the input features (e.g., memory effect and non-linearity) at a reduced network scale. Moreover, a recurrent layer is then utilized to help in learning the temporal behavior of the input signal from the output feature map of the convolutional layer. In the proposed HCRDNN, a dense layer is added after the convolutional and recurrent layers to build a deeper NN model with low computational complexity. Working with hybrid-layers NN architectures enables adjusting the hidden layers' settings to achieve a certain cancellation performance with a considerable computational

complexity reduction.

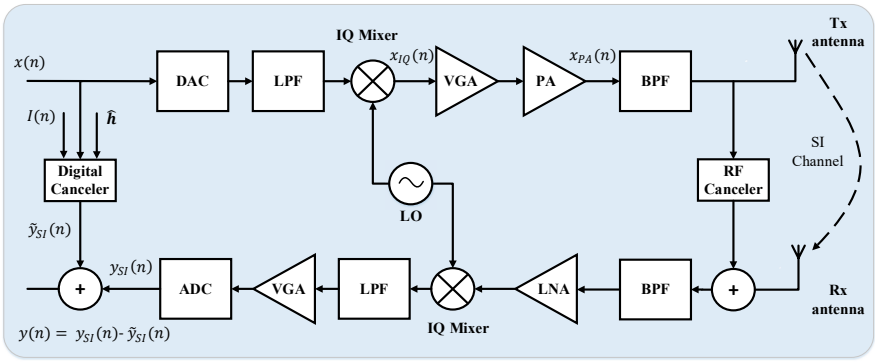
The contributions of this chapter are summarized as follows:

- Two novel hybrid-layers NN architectures, termed as the HCRNN and HCRDNN, are proposed for the first time to model the SI in FD systems with low computational complexity. In contrast to the state-of-the-art NNs that directly apply the traditional dense or recurrent layers for SI modeling, the proposed NNs exploit, in a novel manner, a combination of hidden layers (e.g., convolutional, recurrent, and/or dense) in order to achieve high learning capability while maintaining low computational complexity.

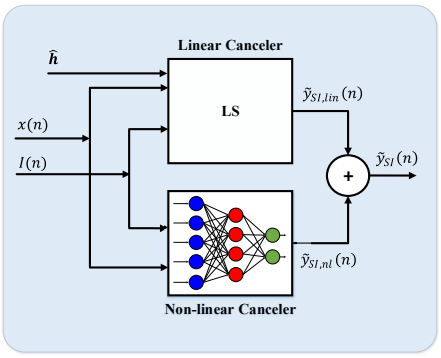
- The computational complexity and memory requirements of the proposed HCRNN and HCRDNN-based cancelers are derived in terms of the number of floating-point operations (FLOPs) and network parameters, respectively, and analyzed compared to those of the polynomial and the state-of-the-art NN-based cancelers.

- The optimum settings for training the proposed HCRNN and HCRDNN architectures (e.g., number of convolutional filters, filter size, number of neurons in recurrent and dense layers, activation functions, learning rate, batch size, and optimizer) are selected to achieve an acceptable cancellation performance with a considerable computational complexity reduction.

- Performance analysis of the two proposed NNs is provided in terms of their prediction capabilities, mean square error (MSE), achieved SI cancellation, computational complexity, and memory requirements. Both NNs demonstrate excellent prediction capabilities in modeling the interference in FD systems with reduced complexity.



(a) Detailed system model.



(b) Digital canceler.

Fig. 3.1: Full-duplex transceiver system model.

The rest of this chapter is organized as follows. Section 3.3 presents the FD transceiver system model. Section 3.4 introduces the proposed HCRNN and HCRDNN-based cancelers. In Section 3.5, the complexity of the proposed NN architectures is analyzed, whereas in Section 3.6, the optimum settings for their training are selected. Finally, simulation results, future research directions, and conclusions are presented in Sections 3.7, 3.8, and 3.9, respectively.

3.3 System Model

An FD transceiver consisting of a local transmitter, local receiver, and two SI cancellation techniques is illustrated in Figs. 3.1(a) and (b). Specifically, the FD system's design, shown in Fig. 3.1(a), employs an analog RF cancellation and a training-based digital cancellation in order to suppress the SI signal to the receiver noise level. The RF cancellation is applied at the first stage of the receiver chain to prevent the SI signal from saturating the receiver's analog components (e.g., LNA, VGA, and ADC). However, the digital cancellation is employed after the ADC to remove the residual SI signal.

Let us denote the digital transmitted samples before the DAC by $x(n)$, with n

representing each sample index. The transmitted samples are converted to analog, filtered, and up-converted to the carrier frequency using the DAC, low pass filter (LPF), and IQ mixer, respectively. The IQ mixer, undesirably, distorts the transmitted signal due to the gain and phase imbalances between the I/Q components (i.e., IQ imbalance). Subsequently, the digital equivalent of the mixer's output signal can be expressed as [16], [29]

$$x_{IQ}(n) = \frac{1}{2}(1 + \psi e^{j\theta})\, x(n) + \frac{1}{2}(1 - \psi e^{j\theta})\, x^*(n), \tag{3.1}$$

where ψ and θ represent the gain and phase imbalance coefficients of the transmitter, respectively. The mixer's output signal is then amplified by the PA, which further distorts the transmitted signal due to its non-idealities. The PA's output signal can be expressed using the conventional parallel-Hammerstein (PH) model, described by (3.48) in the Appendix, as [16], [29], [34]

$$x_{PA}(n) = \sum_{\substack{p=1,\\ p\ odd}}^{P} \sum_{m=0}^{M_{PA}} h_{m,p}\, x_{IQ}(n-m)^{\frac{p+1}{2}} x_{IQ}^*(n-m)^{\frac{p-1}{2}}, \tag{3.2}$$

where $h_{m,p}$ indicates the PA's impulse response. In addition, P and M_{PA} represent the non-linearity order and the PA's memory depth, respectively.

The PA's output signal leaks to the receiver through the SI channel, forming the SI signal. Accordingly, at the receiver side of the FD node, there are three signals: an SI signal, a noise signal, and a far-end desired signal from another FD node. In this work, I assume, for the ease of presentation, that there is no thermal noise, and there are no far-end desired signals from any other FD nodes [29], [32]. As such, the residual SI signal after the RF cancellation process is filtered, amplified, down-converted, and digitized using the band-pass (BPF), LNA, IQ mixer, and ADC, respectively, and can be expressed as [29]

$$y_{_{SI}}(n) = \sum_{\substack{p=1,\\ p\,odd}}^{P} \sum_{q=0}^{p} \sum_{m=0}^{M-1} h_{m,q,p}\, x(n-m)^q x^*(n-m)^{p-q}, \tag{3.3}$$

where $h_{m,q,p}$ represents the impulse response of a channel, including the composite effect of the PA, IQ mixer, and SI channel, whereas M indicates the memory effect incorporated into the input signal by the PA and SI coupling channel.

The estimation of the SI channel is performed using the least-squares (LS) approach as follows. Firstly, I formulate the matrix $\boldsymbol{X}$ with dimensions $N \times M$ from the input data x, where N is the length of the training data. Then, the output signal $\boldsymbol{y}_{_{SI}}$ with dimension $N \times 1$ can be obtained as $\boldsymbol{y}_{_{SI}} = \boldsymbol{X}\boldsymbol{h}$, where $\boldsymbol{h}$ is a vector of size $M \times 1$. In the LS method, the aim is to minimize the cost function $J(\hat{\boldsymbol{h}})$, which can be expressed as

$$\begin{aligned} J(\hat{\boldsymbol{h}}) &= \left\| \boldsymbol{y}_{_{SI}} - \boldsymbol{X}\hat{\boldsymbol{h}} \right\|^2 = (\boldsymbol{y}_{_{SI}} - \boldsymbol{X}\hat{\boldsymbol{h}})^H (\boldsymbol{y}_{_{SI}} - \boldsymbol{X}\hat{\boldsymbol{h}}) \\ &= \boldsymbol{y}_{_{SI}}^H \boldsymbol{y}_{_{SI}} - \boldsymbol{y}_{_{SI}}^H \boldsymbol{X}\hat{\boldsymbol{h}} - \hat{\boldsymbol{h}}^H \boldsymbol{X}^{\boldsymbol{H}} \boldsymbol{y}_{_{SI}} + \hat{\boldsymbol{h}}^H \boldsymbol{X}^{\boldsymbol{H}} \boldsymbol{X}\hat{\boldsymbol{h}}, \end{aligned} \tag{3.4}$$

where $(.)^H$ denotes the conjugate transpose operator. By setting the derivative of this cost function with respect to $\hat{\boldsymbol{h}}$ to zero, the LS solution can be given as

$$\hat{\boldsymbol{h}} = \left(\boldsymbol{X}^{\boldsymbol{H}} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^{\boldsymbol{H}} \boldsymbol{y}_{_{SI}}. \tag{3.5}$$

In the digital canceler, the goal is to estimate the distortion caused by the imperfection of the transceiver hardware components and SI channel to generate an accurate replica of the SI signal $\tilde{y}_{_{SI}}(n)$ at the receiver. This is attained by feeding the baseband transmitted samples before the digital-to-analog conversion to a trainable-based digital canceler in order to produce such a replica. This replica is then subtracted from the SI signal after the ADC to remove the interference, and the residual SI after the

digital cancellation is given by $y(n) = y_{SI}(n) - \tilde{y}_{SI}(n)$. The achieved SI cancellation in the digital domain can be quantified in dB as

$$\mathcal{C}_{dB} = 10 \log_{10} \left(\frac{\sum_n |y_{SI}(n)|^2}{\sum_n |y(n)|^2} \right). \tag{3.6}$$

In this work, the digital canceler is formed by linear and non-linear trainable-based cancelers, as depicted in Fig. 3.1(b). The former is utilized to estimate the linear part of the SI signal based on the conventional LS channel estimation [29], whereas the latter is used to mimic the non-linear part of the SI signal using an NN model. The SI signal is then reconstructed by combining the linear and non-linear components as follows:

$$\tilde{y}_{SI}(n) = \tilde{y}_{SI,lin}(n) + \tilde{y}_{SI,nl}(n), \tag{3.7}$$

where $\tilde{y}_{SI,lin}(n)$ is the linear part of the SI signal, which can be obtained by substituting $p = 1$ and $q = 1$ in (3.3) as

$$\tilde{y}_{SI,lin}(n) = \sum_{m=0}^{M-1} h_{m,1,1}\, x(n-m), \tag{3.8}$$

while $\tilde{y}_{SI,nl}(n)$ is the non-linear part, which can be given as

$$\tilde{y}_{SI,nl}(n) = g\left\{x\left(n\right), I\left(n\right)\right\}, \tag{3.9}$$

where $g\{.\}$ represents the NN mapping function, and $I\left(n\right)$ can be expressed as $I\left(n\right) = \{x\left(n-1\right), x\left(n-2\right), ..., x\left(n-M+1\right)\}$.

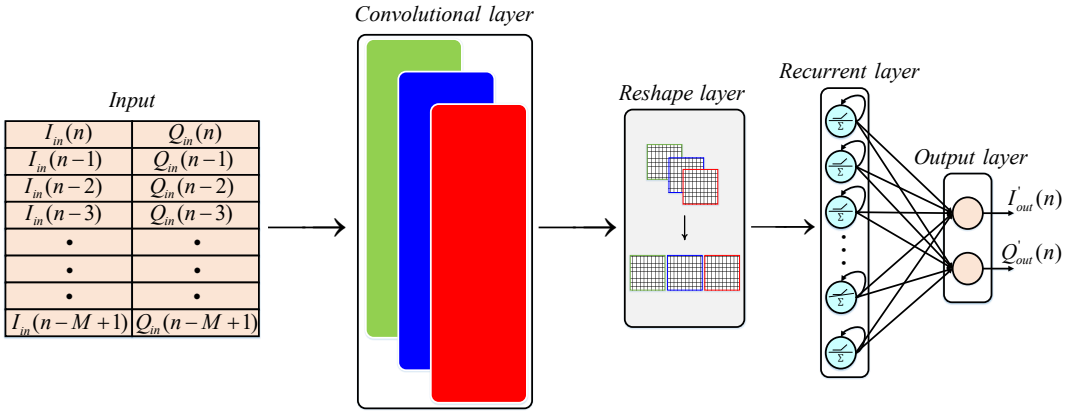


Fig. 3.2: Proposed HCRNN architecture.

3.4 Proposed Hybrid-Layers NN Architectures

The main function of the NN-based canceler is to provide an accurate behavior model to mimic the non-linearities and memory effect attributed to the input signal. As such, to account for such effects, TDNN or RNN-based models were employed in the literature. While the aforementioned models attain good modeling capabilities for approximating the SI in FD systems, their complexity is a considerable issue that should be addressed. A good choice for designing a low-complexity NN model is to exploit the high feature extraction along with the parameter reduction capabilities of the convolutional layer. Further, adding a recurrent layer with a sufficient number of neurons after the convolutional layer can enhance the modeling capabilities with no much increase in the computational complexity. Lastly, exploiting an additional dense layer after the recurrent layer can help to add another degree of freedom for adapting the NN settings in order to achieve the best compromise between the model performance and the computational complexity. In the next subsections, I will show how the proposed NN architectures exploit the aforementioned layers in order to build low-complexity NN-based cancelers.

3.4.1 HCRNN Architecture

The proposed HCRNN architecture is shown in Fig. 3.2. In the HCRNN, hybrid hidden layers are employed to detect the memory effect and non-linearity incorporated into the input signal due to the impairments of the FD transceiver and SI channel. The HCRNN is arranged in five layers: the input layer, convolutional layer, reshape layer, recurrent layer, and output layer. At the input layer, the input data is formulated into a 2D graph consisting of the I/Q components of the instantaneous and delayed versions of the input samples. Arranging the data into a 2D graph allows the input samples to be in an appropriate form that can be efficiently processed by the next convolutional layer. The convolutional layer, the first hidden layer of the HCRNN, is applied to the 2D grid in order to detect the input signal's features with a reduced network scale. Applying the convolutional layer to the input graph comes with a considerable reduction in the computational complexity due to the weight-sharing characteristics and dimensionality reduction capabilities of the convolutional filters [25]. The convolutional layer's output is then reordered using a reshape layer, to be processed by the next recurrent layer, which constitutes the second hidden layer of the HCRNN. The aim of the recurrent layer is to assist in learning the temporal behavior of the input signal from the localized feature map of the convolutional layer. Finally, at the output layer of the HCRNN, the I/Q components of the SI signal are estimated.

An example showing the basic operation of the convolutional and reshape layers of the proposed HCRNN, using $M = 13$, three convolutional filters, and $8 \times 1 \times 1$ filter size, is illustrated in Fig. 3.3. At the convolutional layer, the input data is convolved with the filters, acting as optimizable-feature extractors, in order to detect the patterns of the input attributes. The output feature map of the l^{th} filter after applying the convolution operation can be expressed as

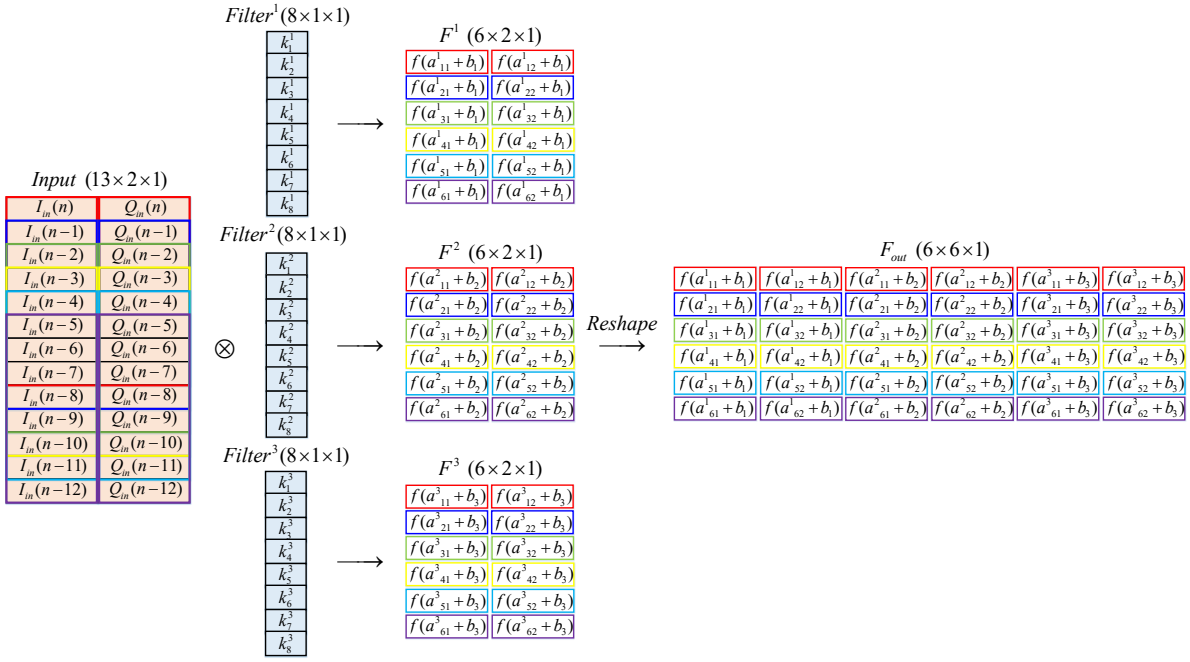


Fig. 3.3: Example of the basic operation of the convolutional and reshape layers of the proposed HCRNN.

$$\boldsymbol{F}_c^l = \begin{bmatrix} I_{in}(n) & Q_{in}(n) \\ I_{in}(n-1) & Q_{in}(n-1) \\ I_{in}(n-2) & Q_{in}(n-2) \\ . & . \\ . & . \\ I_{in}(n-M+1) & Q_{in}(n-M+1) \end{bmatrix} \otimes \begin{bmatrix} k_{1,1}^l & k_{1,S}^l \\ k_{2,1}^l & k_{2,S}^l \\ k_{3,1}^l & k_{3,S}^l \\ . & . \\ . & . \\ k_{R,1}^l & k_{R,S}^l \end{bmatrix}, \tag{3.10}$$

where $l = 1, 2, ..., L$, with L denoting the number of convolutional filters. $I_{in}(n)$ and $Q_{in}(n)$ represent the I/Q components of the current sample $x(n)$, respectively, while $[I_{in}(n-1), ..., I_{in}(n-M+1)]$ and $[Q_{in}(n-1), ..., Q_{in}(n-M+1)]$ indicate the I/Q components of the delayed samples, respectively. In addition, $R \times S \times Z$ and $k_{i,j}^l$ represent the dimensions and the (i^{th}, j^{th}) entry of the l^{th} convolutional filter, respectively. Finally, $\otimes$ denotes the convolution operation. It is worth mentioning

that for all convolutional filters, $R \in \{1, 2, ..., M\}$, $S \in \{1, 2\}$, and $Z = 1$ since the input data is arranged into a 2D graph consisting of two columns to represent the I/Q components of the instantaneous and delayed versions of the input samples.

After performing the convolution, a bias term is added, and a non-linear activation function is applied to each element of $\boldsymbol{F}_c^l$. Accordingly, the output feature map of the l^{th} convolutional filter can be expressed as

$$\boldsymbol{F}^l = \begin{bmatrix} f^c(a_{1,1}^l + b_l) & f^c(a_{1,C}^l + b_l) \\ f^c(a_{2,1}^l + b_l) & f^c(a_{2,C}^l + b_l) \\ f^c(a_{3,1}^l + b_l) & f^c(a_{3,C}^l + b_l) \\ . & . \\ . & . \\ f^c(a_{B,1}^l + b_l) & f^c(a_{B,C}^l + b_l) \end{bmatrix}, \tag{3.11}$$

where $f^c(.)$ denotes the convolutional layer's activation function, while b_l represents the bias term associated with the l^{th} convolutional filter. $B \times C$ indicates the output feature map's dimensions, whereas $a_{i,j}^l$ denotes the (i^{th}, j^{th}) entry of the feature map just after applying the convolution. With unity-stride and without zero-padding, $a_{i,j}^l$ can be given as

$$a_{i,j}^l = \sum_{r=0}^{R-1} \sum_{s=0}^{S-1} \boldsymbol{X}_{r+i,s+j} \boldsymbol{K}_{r+1,s+1}^l, \tag{3.12}$$

where $\boldsymbol{X}$ and $\boldsymbol{K}^l$ represent the input and the l^{th} convolutional filter matrices with entries described in (3.10), respectively.

For efficient processing of the forwarded data through the network, the output feature maps of all filters are reshaped before they are passed to the recurrent layer. More specifically, in the output feature map of each filter, there are dependencies between each column's elements due to the temporal behavior existing in the input

data. Therefore, in the proposed HCRNN, the output feature maps of all filters are reformulated using the reshape layer in order to take these dependencies into account and pass them in a sequence to the recurrent layer. Reshaping the feature maps with this mechanism enables the recurrent layer to detect the aforementioned sequence for proper modeling of the system's temporal behavior. Based on this, the resultant feature map after the reshaping process can be expressed as

$$\boldsymbol{F}_{out} = \begin{bmatrix} \boldsymbol{F}^1 & \boldsymbol{F}^2 & \dots & \boldsymbol{F}^L \end{bmatrix}, \tag{3.13}$$

where $\boldsymbol{F}^1$, $\boldsymbol{F}^2$, and $\boldsymbol{F}^L$ represent the output feature maps of the 1^{st}, 2^{nd}, and L^{th} convolutional filters, respectively. The reshaped feature map is then passed to the recurrent layer, and the output at any time step t can be expressed as

$$\boldsymbol{y}_r(t) = f^r\left(\boldsymbol{f}_{out}(t)\boldsymbol{W}_x + \boldsymbol{y}_r(t-1)\boldsymbol{W}_y + \boldsymbol{b}_r\right), \tag{3.14}$$

where $\boldsymbol{y}_r(t) \in \mathbb{R}^{1\times n_{hr}}$ represents the recurrent layer output at any time step t, with n_{hr} denoting the number of recurrent layer's neurons. Similarly, $\boldsymbol{y}_r(t-1) \in \mathbb{R}^{1\times n_{hr}}$ indicates the recurrent layer's output at the previous time step $t-1$. $\boldsymbol{f}_{out}(t) \in \mathbb{R}^{1\times n_i}$ represents a row vector of $\boldsymbol{F}_{out}$, which is passed to the recurrent layer at time step t, with n_i as the number of input features. $\boldsymbol{W}_x \in \mathbb{R}^{n_i\times n_{hr}}$ denotes the weight matrix for the connections between the input and hidden units at the current time step. Further, $\boldsymbol{W}_y \in \mathbb{R}^{n_{hr}\times n_{hr}}$ indicates the weight matrix for the feedback connections from the hidden units at the previous time step. Finally, $f^r(.)$ represents the activation function operation of the recurrent layer, whereas $\boldsymbol{b}_r \in \mathbb{R}^{1\times n_{hr}}$ indicates a row vector containing the bias terms of the recurrent layer neurons.

The recurrent layer's output is then passed to the output layer, which is formed by a fully-connected (dense) layer containing two neurons. The output layer neurons

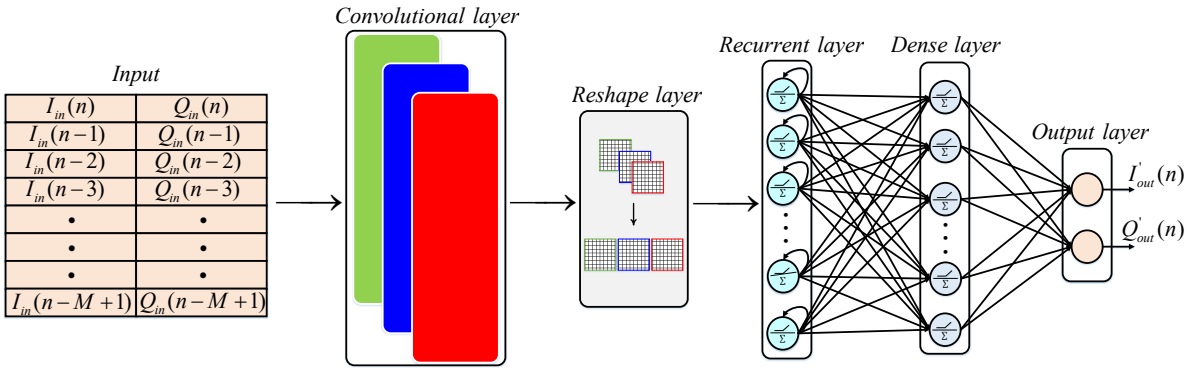


Fig. 3.4: Proposed HCRDNN architecture.

are utilized to map the extracted features by the recurrent layer to the final output (i.e., estimated I/Q components of the SI signal) as follows:

$$I'_{out}(n) = f^o\left(\sum_{i=1}^{n_{hr}} \boldsymbol{W}_o^{1,i}\boldsymbol{y}_r^i + \boldsymbol{b}_o^1\right), \tag{3.15}$$

$$Q'_{out}(n) = f^o\left(\sum_{i=1}^{n_{hr}} \boldsymbol{W}_o^{2,i}\boldsymbol{y}_r^i + \boldsymbol{b}_o^2\right), \tag{3.16}$$

where $f^o(.)$ represents the output layer's activation function, while $\{\boldsymbol{W}_o^{1,i}, \boldsymbol{b}_{o,}^1\}$ and $\{\boldsymbol{W}_o^{2,i}, \boldsymbol{b}_o^2\}$ indicate the weight and bias terms associated with the first and second neurons of the output layer, respectively.

3.4.2 HCRDNN Architecture

The network architecture of the proposed HCRDNN is depicted in Fig. 3.4. The key difference between the HCRDNN and HCRNN architectures is that an additional dense layer is employed after the convolutional and recurrent layers as increasing the number of hidden layers can enable building more complex NN models with reduced complexity [25]. Additionally, adding a dense layer to the HCRDNN architecture provides another degree of freedom to adapt the NN settings in order to achieve

the optimum cancellation-complexity trade-off. In particular, in the HCRDNN, the convolutional layer's hyper-parameters (e.g., number of filters, filter size) and the number of neurons in the recurrent and dense layers are jointly adjusted to achieve a considerable cancellation performance with a significant computational complexity reduction.

The output of the u^{th} neuron at the dense layer of the HCRDNN can be given by

$$y_d^u = f^d\left(\sum_{i=1}^{n_{hr}} \boldsymbol{W}_d^{u,i}\boldsymbol{y}_r^i + \boldsymbol{b}_d^u\right), \tag{3.17}$$

where $f^d(.)$ indicates the activation function of the dense layer, while $\{\boldsymbol{W}_d^{u,i}, \boldsymbol{b}_{d,}^u\}$ represent the weight and bias terms associated with the u^{th} neuron in the dense layer, respectively. The dense layer's output is then passed to the output layer to estimate the I/Q components of the SI signal as (3.15) and (3.16).

After predicting $I_{out}^{'}(n)$ and $Q_{out}^{'}(n)$, the MSE is calculated as a cost function to measure how well the proposed NN models predict the actual outputs as follows:

$$E = \frac{1}{2N}\sum_{n=1}^{N}\left(\Big[I_{out}(n) - I_{out}^{'}(n)\Big]^2 + \Big[Q_{out}(n) - Q_{out}^{'}(n)\Big]^2\right), \tag{3.18}$$

where $\{I_{out}(n), I_{out}(n)\}$ and $\{I_{out}^{'}(n), Q_{out}^{'}(n)\}$ represent the actual and predicted values of the I/Q components of the SI signal, receptively, whereas N denotes the number of training observations, as stated before in Section 3.3. During the HCRNN and HCRDNN models' training, the convolutional, recurrent, and dense layers' weights and biases are set to minimize this cost function.

After the training process, the proposed NNs are employed to provide non-linear SI cancellation as part of the digital cancellation in the FD transceiver, as illustrated in Fig. 3.1(b).

3.5 Complexity Analysis

In this section, the computational complexity and memory requirements of the polynomial and proposed NN-based cancelers are analyzed in terms of the number of computations and memory storage of the linear and non-linear cancelers required to provide the total SI cancellation (i.e., the summation of linear and non-linear cancellations). In particular, in this work, the computational complexity is assessed in terms of the number of FLOPs used to perform the linear and non-linear cancellation processes. Moreover, the memory storage requirements are assessed in terms of the number of stored parameters utilized to achieve the total cancellation. In this analysis, I focus on evaluating the complexity of the polynomial and proposed NN-based cancelers in the real-time inference stage since powerful processing units can be employed to perform the training process offline.

3.5.1 Linear Canceler Complexity

In this subsection, I calculate the number of FLOPs required to implement the linear canceler in terms of the number of RV multiplications and additions. From (3.8), the linear canceler requires M complex multiplications and $M-1$ complex additions to perform the linear cancellation. Using the assumption that each complex addition is implemented by two real additions, and each complex multiplication is executed using three real multiplications and five real additions [32], the number FLOPs of the linear canceler can be expressed as

$$\mathcal{F}_{lin} = \underbrace{3M}_{\chi^{\Re}_{mul,lin}} + \underbrace{7M-2}_{\chi^{\Re}_{add,lin}}, \tag{3.19}$$

where $\chi^{\Re}_{mul,lin}$ and $\chi^{\Re}_{add,lin}$ represent the number of real multiplications and additions required for the linear canceler, respectively, while M denotes the memory effect incorporated into the input signal by the PA and SI channel, as stated before. Similarly, from (3.8), the memory requirements of the linear canceler is evaluated as a function of the number of stored parameters, which can be expressed as

$$\mathcal{P}_{lin} = 2M. \tag{3.20}$$

3.5.2 Non-Linear Canceler Complexity

In this subsection, the number of FLOPs and parameters required for non-linear cancellation using polynomial and NN-based cancelers are analyzed.

3.5.2.1 Polynomial-based Canceler Complexity

Using the previous assumptions, it can be deduced from (3.3) that the number of RV multiplications and additions required for the non-linear polynomial-based canceler can be expressed as [29], [32]

$$\chi^{\Re}_{mul,poly} = 3M \left\{ \left(\frac{P+1}{2} \right) \left(\frac{P+1}{2} + 1 \right) - 1 \right\}, \tag{3.21}$$

$$\chi^{\Re}_{add,poly} = 7M \left\{ \left(\frac{P+1}{2} \right) \left(\frac{P+1}{2} + 1 \right) - 1 \right\}. \tag{3.22}$$

Similarly, the number of parameters of the non-linear polynomial-based canceler can be expressed as [29], [32]

$$\mathcal{P}_{poly} = 2M \left\{ \left(\frac{P+1}{2} \right) \left(\frac{P+1}{2} + 1 \right) - 1 \right\}. \tag{3.23}$$

3.5.2.2 NN-based Canceler Complexity

For the proposed non-linear HCRNN-based canceler, the number of FLOPs can be expressed as

$$\mathcal{F}_{HCRNN} = \mathcal{F}^c + \mathcal{F}^r + \mathcal{F}^o, \tag{3.24}$$

where $\mathcal{F}^c$, $\mathcal{F}^r$, and $\mathcal{F}^o$ represent the number of FLOPs required for the convolutional, recurrent, and output layers of the HCRNN, respectively. Firstly, for the convolutional layer, $\mathcal{F}^c$ can be expressed as

$$\mathcal{F}^c = \underbrace{\left(2\left\{R \times S \times Z\right\} - 1\right)\left(B \times C \times L\right)}_{\text{Term}_1^c} + \underbrace{\mathcal{F}^{act} \times \left(B \times C \times L\right)}_{\text{Term}_2^c} + \underbrace{\left(B \times C \times L\right)}_{\text{Term}_3^c}, \tag{3.25}$$

where Term_1^c represents the number of RV multiplications and additions required to convolve the 2D graph of the input data with L convolutional filters. Term_2^c denotes the number of RV operations (multiplications + additions) required for applying the activation functions at each element of the output feature map after the convolution process. Term_3^c indicates the number of real additions required for adding the bias values. Finally, $\mathcal{F}^{act}$ represents the number of real operations needed to evaluate each activation function, which mainly depends on the activation function type.

In this work, I consider the rectified linear unit ($ReLU$), $Sigmoid$, and hyperbolic

tangent ($tanh$) activation functions, which are defined as [25]

$$ReLU(z) = \max(0, z), \tag{3.26a}$$

$$Sigmoid(z) = \frac{1}{e^{-z}+1}, \tag{3.26b}$$

$$tanh(z) = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}}. \tag{3.26c}$$

By assuming that each real operation (e.g., multiplication, division, addition, subtraction, and exponentiation) costs one FLOP [35], [36], the number of FLOPs required to implement each of the aforementioned activation functions can be expressed as

$$\mathcal{F}^{act} = \begin{cases} 1, & \text{if } ReLU \\ 4, & \text{if } Sigmoid \cdot \\ 6, & \text{if } tanh \end{cases} \tag{3.27}$$

Secondly, the number of FLOPs required for the recurrent layer can be expressed as

$$\mathcal{F}^{r} = \underbrace{2n_{hr}\left(n_i + n_{hr} - \frac{1}{2}\right)}_{\text{Term}_1^r} + \underbrace{\mathcal{F}^{act} \times n_{hr}}_{\text{Term}_2^r} + \underbrace{n_{hr}}_{\text{Term}_3^r}, \tag{3.28}$$

where Term_1^r represents the number of RV operations associated with multiplying the weight matrices $\boldsymbol{W}_x$ and $\boldsymbol{W}_y$ in (3.14) with their corresponding inputs, while Term_2^r and Term_3^r indicate the number of real operations required for applying the activation functions and adding the biases at each neuron of the recurrent layer, respectively.

Finally, the output layer's FLOPs can be expressed as

$$\mathcal{F}^o = \underbrace{n_{ho}\left(2n_{hr}-1\right)}_{\text{Term}_1^o} + \underbrace{\mathcal{F}^{act} \times n_{ho}}_{\text{Term}_2^o} + \underbrace{n_{ho}}_{\text{Term}_3^o}, \tag{3.29}$$

where n_{ho} represents the number of output layer's neurons. Term_1^o denotes the number of RV operations required to calculate the weighted sum of inputs at the output layer neurons. Term_2^o and Term_3^o indicate the number of real operations associated with employing the activation functions and adding the bias terms at each neuron of the output layer, respectively.

The number of parameters of the proposed HCRNN architecture can be given as

$$\mathcal{P}_{HCRNN} = \mathcal{P}^c + \mathcal{P}^r + \mathcal{P}^o, \tag{3.30}$$

where $\mathcal{P}^c$, $\mathcal{P}^r$, and $\mathcal{P}^o$ represent the number of parameters of the convolutional, recurrent, and output layers, respectively, which can be expressed as

$$\mathcal{P}^c = L\left(R \times S \times Z + 1\right), \tag{3.31a}$$

$$\mathcal{P}^r = n_{hr}\left(n_i + n_{hr} + 1\right), \tag{3.31b}$$

$$\mathcal{P}^o = n_{ho}\left(n_{hr} + 1\right). \tag{3.31c}$$

Using the same mathematical formulation, the number of FLOPs of the proposed HCRDNN can be calculated as

$$\mathcal{F}_{HCRDNN} = \mathcal{F}^c + \mathcal{F}^r + \mathcal{F}^d + \mathcal{F}^o, \tag{3.32}$$

where $\mathcal{F}^c$ and $\mathcal{F}^r$ represent the convolutional and recurrent layers' FLOPs, which can be determined using (3.25) and (3.28), respectively. $\mathcal{F}^o$ indicates the output layer's

FLOPs, which can be calculated by replacing n_{hr} by n_{hd} in (3.29), whereas $\mathcal{F}^d$ denotes the dense layer's FLOPs, which can be expressed as

$$\mathcal{F}^d = \underbrace{n_{hd}\left(2n_{hr}-1\right)}_{\text{Term}_1^d} + \underbrace{\mathcal{F}^{act}\times n_{hd}}_{\text{Term}_2^d} + \underbrace{n_{hd}}_{\text{Term}_3^d}, \tag{3.33}$$

where Term_1^d, Term_2^d, and Term_3^d represent the number of real operations required for calculating the weighted sum of inputs, applying the activation functions, and summing the biases at the dense layer neurons, respectively.

The number of parameters of the HCRDNN can be given as

$$\mathcal{P}_{HCRDNN} = \mathcal{P}^c + \mathcal{P}^r + \mathcal{P}^d + \mathcal{P}^o, \tag{3.34}$$

where $\mathcal{P}^c$ and $\mathcal{P}^r$ represent the convolutional and recurrent layers' parameters given by (3.31a) and (3.31b), respectively, while $\mathcal{P}^o$ indicates the output layer's parameters, which can be evaluated by replacing n_{hr} by n_{hd} in (3.31c). Lastly, $\mathcal{P}^d = n_{hd}\left(n_{hr}+1\right)$ is the number of dense layer parameters.

For the RV-TDNN [29], [30], the number of FLOPs and parameters can be expressed as

$$\mathcal{F}_{RV-TDNN} = \sum_{j=2}^{\mathcal{L}} 2n_{j-1}n_j + \mathcal{F}^{act,j}n_j, \tag{3.35}$$

$$\mathcal{P}_{RV-TDNN} = \sum_{j=2}^{\mathcal{L}} n_{j-1}n_j + n_j, \tag{3.36}$$

where $\mathcal{L}$ indicates the number of layers of the RV-TDNN, including the input, hidden, and output layers. n_j is the number of neurons in the j^{th} layer, with n_1 and $n_{\mathcal{L}}$ representing the number of input and output layers' neurons, respectively. $\mathcal{F}^{act,j}$ denotes the number of operations required to apply the activation function at each of

the j^{th} layer's neurons.

Similarly, for the RNN [30], the number of FLOPs and parameters can be given by

$$\mathcal{F}_{RNN} = \sum_{j=2}^{\mathcal{L}} 2n_{j-1}n_j + 2n_j^2 + \mathcal{F}^{act,j}n_j - 2n_{\mathcal{L}}^2, \tag{3.37}$$

$$\mathcal{P}_{RNN} = \sum_{j=2}^{\mathcal{L}} n_{j-1}n_j + n_j^2 + n_j - n_{\mathcal{L}}^2, \tag{3.38}$$

with n_1 denotes the number of RNN inputs at each time step.

For the CV-TDNN [30], using the previous implementation assumptions for the CV additions and multiplications, the number of FLOPs and parameters can be given by

$$\mathcal{F}_{CV-TDNN} = 10\left\{\sum_{j=2}^{\mathcal{L}} n_{j-1}n_j\right\} + \mathcal{F}^{act,j}n_j, \tag{3.39}$$

$$\mathcal{P}_{CV-TDNN} = 2\left\{\sum_{j=2}^{\mathcal{L}} n_{j-1}n_j + n_j\right\}. \tag{3.40}$$

For the LWGS [33] employing n_i CV inputs and a single hidden layer with n_h neurons, the number of FLOPs and parameters can be expressed as

$$\mathcal{F}_{LWGS} = 10\left\{\left(\sum_{j=1}^{n_h} j\right) + n_i\right\} + \mathcal{F}^{act}n_h, \tag{3.41}$$

$$\mathcal{P}_{LWGS} = 2\left\{\left(\sum_{j=1}^{n_h} j\right) + n_i + n_h{+}1\right\}. \tag{3.42}$$

Likewise, for the MWGS [33] employing n_i CV inputs, a $\mathcal{W}$ window size, and a single hidden layer with n_h neurons, the number of FLOPs and parameters can be obtained

as follows:

$$\mathcal{F}_{MWGS} = 10\left\{n_i + \mathcal{W}\left(n_h - 1\right) + n_h\right\} + \mathcal{F}^{act} n_h, \tag{3.43}$$

$$\mathcal{P}_{MWGS} = 2\left\{n_i + \mathcal{W}\left(n_h - 1\right) + 2n_h + 1\right\}. \tag{3.44}$$

Finally, the total number of FLOPs of any of the aforementioned NN-based cancelers can be expressed as

$$\mathcal{F}_{NN-\mathcal{C}} = \mathcal{F}_{lin} + \mathcal{F}_{NN} + 2, \tag{3.45}$$

where $\mathcal{F}_{lin}$ represents the linear canceler complexity, which can be calculated by (3.19), while $\mathcal{F}_{NN}$ indicates the NN complexity (e.g., $\mathcal{F}_{HCRNN}$, $\mathcal{F}_{HCRDNN}$, $\mathcal{F}_{RV-TDNN}$). The two FLOPs added in (3.45) represent the number of real additions required to sum the outputs of the linear and non-linear cancelers in (3.8) and (3.9), respectively [32].

3.6 Optimum Settings of the Proposed HCRNN and HCRDNN Architectures

In this section, the optimum settings for training the proposed HCRNN and HCRDNN architectures are obtained. Specifically, the number of convolutional filters, filter size, number of recurrent and dense layers' neurons, activation functions in each layer, learning rate, batch size, and training optimizer are selected in such a way that results in a proper cancellation performance while maintaining low computational complexity. In the following, I firstly describe the dataset employed to train the proposed NN architectures, then the network settings for the training process are optimized.

TABLE 3.1: Summary of the parameters employed to generate the dataset utilized for training and verifying the proposed NN architectures.

Parameter	Value
Type of modulation	QPSK-modulated OFDM
Pass-band bandwidth	10 MHz
Number of carriers	1024
Sampling frequency	20 MHz
Average transmit power	10 dBm
Passive analog suppression	53 dB
Active analog suppression	N/A
Total analog cancellation	53 dB
Dataset size	20,480 samples

3.6.1 Training Dataset

In the following experiments, the public dataset employed in [29] and [30] is utilized to train and verify the proposed NN architectures. The dataset is produced using a realistic FD test-bed, which generates a 10 MHz quadrature phase-shift keying modulated-orthogonal frequency division multiplexing (OFDM) signal with an average transmit power of 10 dBm. The OFDM signal employs 1024 sub-carriers and is sampled at 20 MHz. The generated dataset contains 20,480 time-domain baseband samples, which I split into two distinct parts. The first part is used for training and consists of 90% of samples, whereas the second is employed for verification and contains the rest 10%. The employed hardware test-bed provides 53 dB passive analog RF cancellation using physical separation between the transmit and receive antennas; hence, in this work, no further active analog cancellation techniques are employed since the use of digital cancellation after passive analog suppression is adequate to bring the SI signal's power down to the receiver's noise level [29], [30]. A summary of the parameters employed to generate the aforementioned dataset is presented in Table 3.1.

3.6.2 Optimum Configuration

The process of selecting the optimum configuration of the proposed HCRNN (e.g., number of convolutional filters, filter size, number of recurrent layer's neurons) is analyzed as given in Table 3.2.[2] To that end, I test two values for the convolutional filters $L \in \{2, 3\}$ and change the filter size dimensions such that $R \in \{2, 3, ..., 13\}$, $S \in \{1, 2\}$, and $Z = 1$ for all filters. Moreover, I consider $n_{hr} \in \{4, 5, ..., 10\}$ for the number of neurons in the recurrent layer. Since there is a large number of combinations between L, R, S, Z, and n_{hr}, I only show the NN configurations that achieve the best cancellation-complexity trade-off in Table 3.2. Specifically, in this work, choosing the optimum configuration of the HCRNN is based on the criterion of achieving a similar or comparable cancellation performance to the polynomial canceler with $P = 5$ while maintaining a lower implementation complexity.[3] Based on this, it can be observed from Table 3.2 that the HCRNN configuration with three convolutional filters, $12 \times 1 \times 1$ filter size, and nine neurons in the recurrent layer achieves the target cancellation of the polynomial canceler. Besides, it attains the best compromise between the cancellation performance and implementation complexity compared to its other counterparts.

Similarly, the selection of the optimum configuration of the proposed HCRDNN is explored in Table 3.3.[2] In this case, I test the number of filters $L \in \{2, 3\}$, filter dimensions $R \in \{2, 3, ..., 13\}$, $S \in \{1, 2\}$, and $Z = 1$, recurrent layer's neurons $n_{hr} \in \{4, 5, ..., 10\}$, and dense layer's neurons $n_{hd} \in \{4, 5, ..., 12\}$. As can be seen from Table 3.3, two configurations of the proposed HCRDNN achieve the best compromise

[2]The results in Tables 3.2 and 3.3 are obtained using the following default settings: *ReLU* activation function in all hidden layers, $5{\times}10^{-3}$ learning rate, 158 batch size, and Adam optimization algorithm over 15 random initializations. The aforementioned settings are optimized in the next subsections.

[3]It is noted that at $P = 5$, the polynomial canceler attains 44.45 dB cancellation [30], while from (3.19), (3.20), (3.21), (3.22), and (3.23), it requires 1558 FLOPs and 312 parameters to achieve this cancellation.

TABLE 3.2: Selection of the optimum configuration of the proposed HCRNN.

Config. #	**# Filters**	**Filter size**	**# Rec. neurons**	**SI Canc. (dB)**	**Complexity**		**Config. #**	**# Filters**	**Filter size**	**# Rec. neurons**	**SI Canc. (dB)**	**Complexity**	
					# Par.	# FLOPs						# Par.	# FLOPs
1	2	4×1×1	6	43.01	116	640	17	2	4×1×1	10	44.18	208	820
2	2	8×1×1	6	42.60	124	688	18	2	9×1×1	10	44.05	218	840
3	2	6×1×1	7	42.92	140	735	19	2	10×1×1	10	44.06	220	796
4	2	4×1×1	8	43.75	158	722	20	2	11×1×1	10	44.19	222	736
5	2	5×1×1	8	43.81	160	758	21	2	12×1×1	10	44.31	224	660
6	2	8×1×1	8	43.89	166	770	22	3	11×1×1	6	43.66	154	718
7	2	10×1×1	8	43.89	170	698	23	3	12×1×1	6	43.97	157	604
8	2	11×1×1	8	43.68	172	638	24	3	11×1×1	7	44.11	176	761
9	2	5×1×1	9	44.03	184	805	25	3	12×2×1	7	43.54	194	599
10	2	6×1×1	9	44.03	186	825	26	3	11×1×1	8	44.10	200	808
11	2	7×1×1	9	44.02	188	829	**27**	**3**	**12×1×1**	**9**	**44.44**	**229**	**745**
12	2	9×1×1	9	43.93	192	789	28	3	13×1×1	9	43.94	232	607
13	2	10×1×1	9	44.08	194	745	29	3	11×2×1	9	43.85	232	796
14	2	12×1×1	9	44.23	198	609	30	3	12×2×1	9	44.09	238	685
15	2	12×2×1	9	43.61	204	569	31	3	12×1×1	10	44.49	257	800
16	2	3×1×1	10	44.06	206	768	32	3	12×2×1	10	44.12	299	734

TABLE 3.3: Selection of the optimum configuration of the proposed HCRDNN.

Config. #	**# Filters**	**Filter size**	**# Rec. neurons**	**# Dense neurons**	**SI Canc. (dB)**	**Complexity**		**Config. #**	**# Filters**	**Filter size**	**# Rec. neurons**	**# Dense neurons**	**SI Canc. (dB)**	**Complexity**	
						# Par.	# FLOPs							# Par.	# FLOPs
1	2	7×1×1	5	9	44.10	166	789	17	2	12×1×1	8	10	44.42	268	740
2	2	9×1×1	5	10	44.10	178	755	18	2	12×1×1	9	7	44.42	264	734
3	2	12×1×1	5	11	44.29	192	590	19	2	12×1×1	9	8	44.50	276	757
4	2	9×1×1	5	12	44.14	194	785	20	2	12×1×1	9	9	44.42	288	780
5	2	10×1×1	5	12	44.18	196	741	21	2	12×1×1	10	6	44.46	282	770
6	2	12×1×1	5	12	44.27	200	605	22	3	12×1×1	5	6	44.11	175	635
7	2	10×1×1	6	9	44.09	197	745	23	3	12×1×1	5	7	44.10	183	650
8	2	12×1×1	6	9	44.22	201	609	24	3	12×1×1	5	8	44.15	191	665
9	2	11×1×1	6	10	44.11	208	702	25	3	12×1×1	5	10	44.24	207	695
10	2	10×1×1	6	11	44.25	215	779	26	3	12×1×1	5	11	44.17	215	710
11	2	11×1×1	6	11	44.23	217	719	27	**3**	**12×1×1**	**5**	**12**	**44.41**	**223**	**725**
12	2	12×1×1	6	11	44.26	219	643	28	3	12×1×1	6	6	44.28	199	682
13	2	11×1×1	7	11	44.21	246	776	29	3	12×1×1	6	8	44.22	217	716
14	**2**	**12×1×1**	**7**	**11**	**44.44**	**248**	**700**	30	3	12×1×1	6	9	44.36	226	733
15	2	12×1×1	7	12	44.44	258	719	31	3	12×1×1	6	11	44.40	244	767
16	2	12×1×1	8	9	44.41	257	719	32	3	12×1×1	6	12	44.39	253	784

TABLE 3.4: Optimum activation functions of the proposed HCRNN.

Config. #	**Conv. layer**	**Rec. layer**	**SI Canc. (dB)**	**FLOPs**
1	*tanh*	*tanh*	44.61	850
2	*tanh*	*ReLU*	44.50	805
3	*tanh*	*Sigm*	44.15	832
4	*ReLU*	*tanh*	44.37	790
5	***ReLU***	***ReLU***	**44.44**	**745**
6	*ReLU*	*Sigm*	44.41	772
7	*Sigm*	*tanh*	44.27	826
8	*Sigm*	*ReLU*	43.65	781
9	*Sigm*	*Sigm*	43.12	808

among the cancellation performance, number of FLOPs, and parameters. The first configuration (i.e., HCRDNN 1) employs two convolutional filters, $12 \times 1 \times 1$ filter size, seven neurons in the recurrent layer, and eleven neurons in the dense layer. However, the second (i.e., HCRDNN 2) utilizes three convolutional filters, $12 \times 1 \times 1$ filter size, five neurons in the recurrent layer, and twelve neurons in the dense layer. It is worth mentioning here that from Tables 3.2 and 3.3, the filter size of $12 \times 1 \times 1$ is shown to be the best size for the optimizable convolutional filters employed in both the HCRNN and HCRDNN architectures. Moreover, it is worth noting that using an additional dense layer in the proposed HCRDNN 1 and HCRDNN 2 reduces the number of neurons in the recurrent layer from 9 to 7 and 5 neurons, respectively, compared to the HCRNN architecture; this significantly affects the computational complexity of the HCRDNN model.

3.6.3 Optimum Activation Functions

In this subsection, I test the cancellation performance and computational complexity of the proposed HCRNN and HCRDNN using different activation functions in the convolutional, recurrent, and dense (if any) layers and select the optimum combination of activation functions that results in the best cancellation-complexity trade-off.

TABLE 3.5: Optimum activation functions of the proposed HCRDNN.

Config.	**Conv.**	**Rec.**	**Dense**	**SI Canc. (dB)**		**FLOPs**	
#	**layer**	**layer**	**layer**	**HCRDNN 1**	**HCRDNN 2**	**HCRDNN 1**	**HCRDNN 2**
1	*tanh*	*tanh*	*tanh*	44.41	44.33	830	870
2	*tanh*	*tanh*	*ReLU*	44.49	44.57	775	810
3	*tanh*	*tanh*	*Sigm*	44.21	43.98	808	846
4	*tanh*	*ReLU*	*tanh*	44.40	44.30	795	845
5	*tanh*	*ReLU*	*ReLU*	44.36	44.34	740	785
6	*tanh*	*ReLU*	*Sigm*	44.14	43.98	773	821
7	*tanh*	*Sigm*	*tanh*	43.98	44.05	816	860
8	*tanh*	*Sigm*	*ReLU*	44.13	44.34	761	800
9	*tanh*	*Sigm*	*Sigm*	43.96	44.09	794	836
10	*ReLU*	*tanh*	*tanh*	44.48	44.24	790	810
11	*ReLU*	*tanh*	*ReLU*	44.54	44.48	735	750
12	*ReLU*	*tanh*	*Sigm*	44.17	43.88	768	786
13	*ReLU*	*ReLU*	*tanh*	44.36	44.14	755	785
14	***ReLU***	***ReLU***	***ReLU***	**44.44**	**44.41**	**700**	**725**
15	*ReLU*	*ReLU*	*Sigm*	44.24	43.92	733	761
16	*ReLU*	*Sigm*	*tanh*	44.34	43.89	776	800
17	*ReLU*	*Sigm*	*ReLU*	44.31	44.26	721	740
18	*ReLU*	*Sigm*	*Sigm*	44.11	43.95	754	776
19	*Sigm*	*tanh*	*tanh*	44.08	43.89	814	846
20	*Sigm*	*tanh*	*ReLU*	44.04	44.07	759	786
21	*Sigm*	*tanh*	*Sigm*	43.90	43.59	792	822
22	*Sigm*	*ReLU*	*tanh*	43.46	42.78	779	821
23	*Sigm*	*ReLU*	*ReLU*	43.95	43.63	724	761
24	*Sigm*	*ReLU*	*Sigm*	43.84	43.21	757	797
25	*Sigm*	*Sigm*	*tanh*	42.53	41.98	800	836
26	*Sigm*	*Sigm*	*ReLU*	43.02	43.54	745	776
27	*Sigm*	*Sigm*	*Sigm*	42.97	42.94	778	812

TABLE 3.6: Optimum learning rate of the proposed HCRNN and HCRDNN.

Learning rate	$\mathbf{1 \times 10^{-2}}$	$\mathbf{5 \times 10^{-3}}$	$\mathbf{1 \times 10^{-3}}$	$\mathbf{5 \times 10^{-4}}$	$\mathbf{1 \times 10^{-4}}$
Network	**SI Cancellation (dB)**				
HCRNN	44.36	**44.44**	44.12	43.64	38.90
HCRDNN 1	44.28	**44.44**	43.65	42.92	39.11
HCRDNN 2	44.22	**44.41**	44.01	43.26	39.22

Specifically, in Table 3.4, I evaluate the performance of the optimum HCRNN configuration, obtained in the previous subsection, using different activation functions. As seen from Table 3.4, the HCRNN with the *ReLU* activation function in both convolutional and recurrent layers achieves the target cancellation of the polynomial canceler and provides the best compromise between the cancellation and complexity performances. It can also be inferred from Table 3.4 that using the *tanh* activation function in the hidden layers of HCRNN only enhances the SI cancellation by 0.17 dB at the cost of augmenting the FLOPs by 14% compared to the *ReLU* activation function. Hence, in this work, I employ the *ReLU* activation function for the HCRNN as it provides the best cancellation-complexity trade-off.

Likewise, in Table 3.5, I test the optimum configurations of the HCRDNN using various activation functions. As can be observed, using the *ReLU* activation function in the convolutional, recurrent, and dense layers attains the best cancellation-complexity trade-off for both HCRDNN architectures.

3.6.4 Optimum Learning Rate

The effect of varying the learning rate on the cancellation performance of the proposed HCRNN and HCRDNN-based cancelers is analyzed in Table 3.6. It can be inferred that using a learning rate of $5{\times}10^{-3}$ achieves the best cancellation performance for both the proposed HCRNN and HCRDNN architectures compared to the other learning rates.

TABLE 3.7: Optimum batch size of the proposed HCRNN and HCRDNN.

Batch size	**22**	**62**	**158**	**256**	**512**
Network	**SI Cancellation (dB)**				
HCRNN	44.37	**44.50**	44.44	44.40	44.44
HCRDNN 1	44.17	44.38	**44.44**	44.39	44.28
HCRDNN 2	44.22	44.29	**44.41**	44.37	44.31

3.6.5 Optimum Batch Size

Similarly, in this subsection, I test the effect of varying the batch size on the performance of the proposed HCRNN and HCRDNN-based cancelers. Herein, I consider many values for the batch size, and I only show the batch sizes that result in the best cancellation performance in Table 3.7. As can be observed, employing batch size values of 62 and 158 provide the best cancellation performance for the proposed HCRNN and HCRDNN architectures, respectively.

3.6.6 Selected Optimizer

The cancellation performance of the proposed HCRNN and HCRDNN-based cancelers is analyzed using different optimizers, as illustrated in Table 3.8. In this work, I test the stochastic gradient descent (SGD), Adam, root mean square propagation (RMSprop), Adadelta, and Adamax optimizers. As seen from Table 3.8, the Adam optimization algorithm attains the best performance for the HCRNN and HCRDNN compared to the other optimization techniques.

Based on the aforementioned subsections, the optimum settings for training the proposed HCRNN and HCRDNN architectures are summarized in Table 3.9.

TABLE 3.8: Selected optimizer of the proposed HCRNN and HCRDNN.

Optimizer	**SGD**	**Adam**	**RMSprop**	**Adadelta**	**Adamax**
Network	**SI Cancellation (dB)**				
HCRNN	41.11	**44.50**	44.23	37.80	44.36
HCRDNN 1	38.37	**44.44**	43.97	37.84	43.93
HCRDNN 2	38.73	**44.41**	43.89	37.83	43.95

3.7 Results and Discussions

In this section, the proposed and the state-of-the-art NN-based cancelers are assessed for canceling the SI in a realistic FD system and compared with the polynomial-based canceler's performance. The performance assessment includes studying the proposed NN architectures' prediction capabilities, MSE performance, achieved SI cancellation, power spectral density (PSD) of the modeled SI signal, computational complexity, and memory requirements. The above-mentioned metrics are considered the typical measures for performance evaluation in the SI cancellation-related works in the literature [29]-[33]. Herein, for the sake of comprehensive evaluation, I compare the performance of the proposed NNs with the shallow and deep RV-TDNN, RNN, and CV-TDNN introduced in [29] and [30]; further, I consider the performance of the LWGS and MWGS investigated in [33]. In this work, all NN architectures are implemented using 3.5.7 Python software with 2.0.0 TensorFlow and 2.3.1 Keras versions. Moreover, the NNs are trained using the dataset described in Section 3.6.1 over 15 random network initializations. For the proposed NN architectures, I employ the optimum settings summarized in Table 3.9, whereas for the state-of-the-art NNs, I use the optimum settings listed in Table 3.10. Besides, I employ a memory length $M = 13$ for the polynomial and all NN-based cancelers. It is worth mentioning that the optimum settings for training the proposed and the state-of-the-art NN-based cancelers, summarized in Tables 3.9 and 3.10 respectively, are chosen by applying trial and error methodology

TABLE 3.9: Optimum settings of the proposed HCRNN and HCDRNN.

Structure	**HCRNN**	**HCRDNN 1**	**HCRDNN 2**
# Filters	3	2	3
Filter size	12×1×1	12×1×1	12×1×1
# Rec. neurons	9	7	5
# Dense neurons	N/A	11	12
Conv. activation	*ReLU*	*ReLU*	*ReLU*
Rec. activation	*ReLU*	*ReLU*	*ReLU*
Dense activation	N/A	*ReLU*	*ReLU*
Learning rate	5×10^{-3}	5×10^{-3}	5×10^{-3}
Batch size	62	158	158
Optimizer	Adam	Adam	Adam

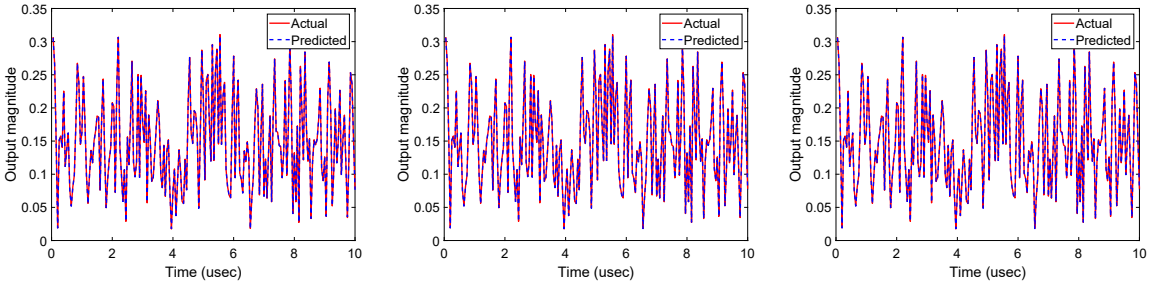
TABLE 3.10: Optimum settings of the state-of-the-art NN architectures.

Structure	**# Hidden layers neurons**	**Activation**	**Learning rate**	**Batch size**	**Opti-mizer**
RV-TDNN [8]	18	*ReLU*	5×10^{-3}	22	Adam
RNN [11]	20	*tanh*	2.5×10^{-3}	158	Adam
CV-TDNN [11]	7	$\mathbb{C}$*ReLU*	4.5×10^{-3}	62	Adam
LWGS [33]	9	$\mathbb{C}$*ReLU*	4.5×10^{-3}	62	Adam
MWGS [33]	12	$\mathbb{C}$*ReLU*	4.5×10^{-3}	62	Adam
Deep RV-TDNN [11]	(10-10-10)	*ReLU*	5×10^{-3}	22	Adam
Deep RNN [11]	(16-16-16)	*tanh*	2.5×10^{-3}	158	Adam
Deep CV-TDNN [11]	(4-4-4)	$\mathbb{C}$*ReLU*	4.5×10^{-3}	62	Adam

such they achieve a similar cancellation performance to the polynomial canceler at $P = 5$ (i.e., 44.45 dB cancellation).

3.7.1 Prediction Capabilities of the Proposed NN Architectures

In this subsection, the prediction capabilities of the proposed HCRNN, HCRDNN 1, and HCRDNN 2 architectures are assessed in Figs. 3.5(a), (b), and (c), respectively. Herein, I show the time-domain waveforms for 200 input-output sample pairs of the SI signal predicted by the proposed NN architectures and their corresponding ground-truth (actual) values. As can be seen from the figures, there is a consistency between



(a) HCRNN. (b) HCRDNN 1. (c) HCRDNN 2.

Fig. 3.5: Prediction capabilities of the proposed NN architectures.

the actual and predicted values of the SI signal modeled by the proposed NNs. This consistency substantiates the ability of the proposed HCRNN and HCRDNN-based cancelers to model the SI correctly. In addition, it can be inferred from the figures that the proposed NNs have similar modeling capabilities since their network settings are intentionally set such that they achieve a comparable cancellation performance to the polynomial-based canceler with $P = 5$.

3.7.2 MSE Performance

The MSE performance in the training and testing phases of the proposed NN architectures is depicted in Figs. 3.6 and 3.7, respectively, compared to the state-of-the-art NNs. The MSE is utilized to evaluate the average squared difference between the ground-truth and predicted SI signal modeled by the various NN architectures. As seen from the figures, the MSE values of all NNs are comparable in both training and testing phases (this can be clearly observed from the inset graphs in Figs. 3.6 and 3.7). As previously stated, the reason behind this is that all NNs are set to attain a comparable cancellation performance, and that is why they achieve a similar MSE performance. Moreover, it can be observed from Figs. 3.6 and 3.7 that there are no overfitting signs for the proposed and the state-of-the-art NNs as they perform well

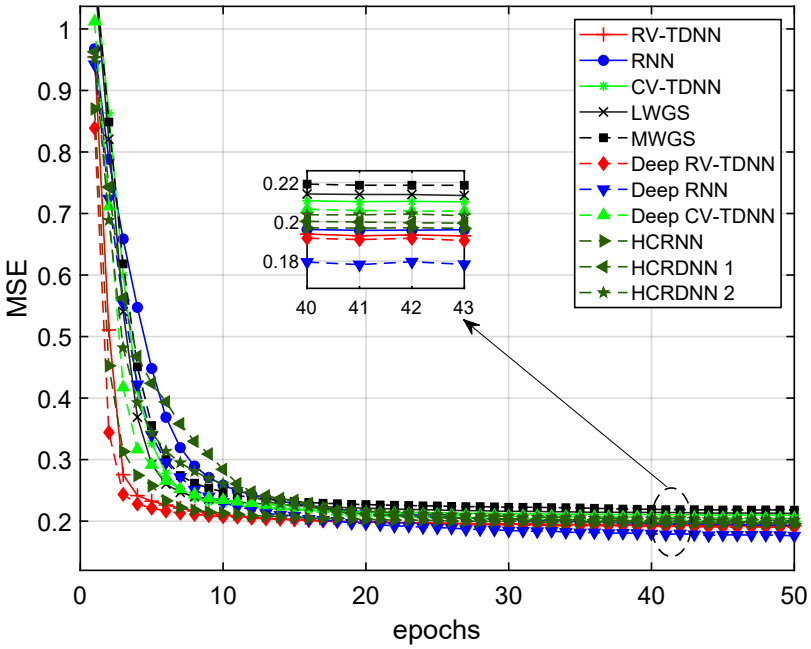


Fig. 3.6: MSE in the training phase for the proposed and the state-of-the-art NN architectures.

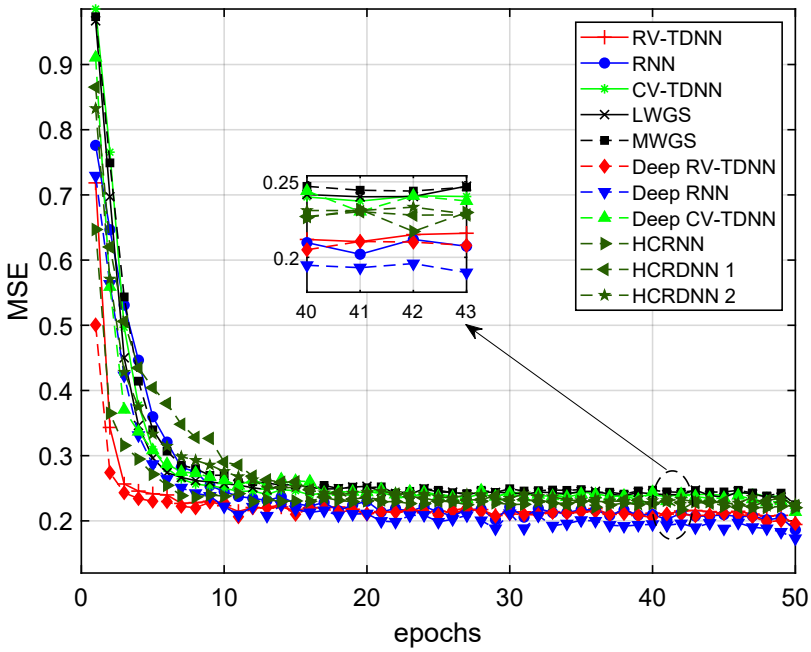


Fig. 3.7: MSE in the testing phase for the proposed and the state-of-the-art NN architectures.

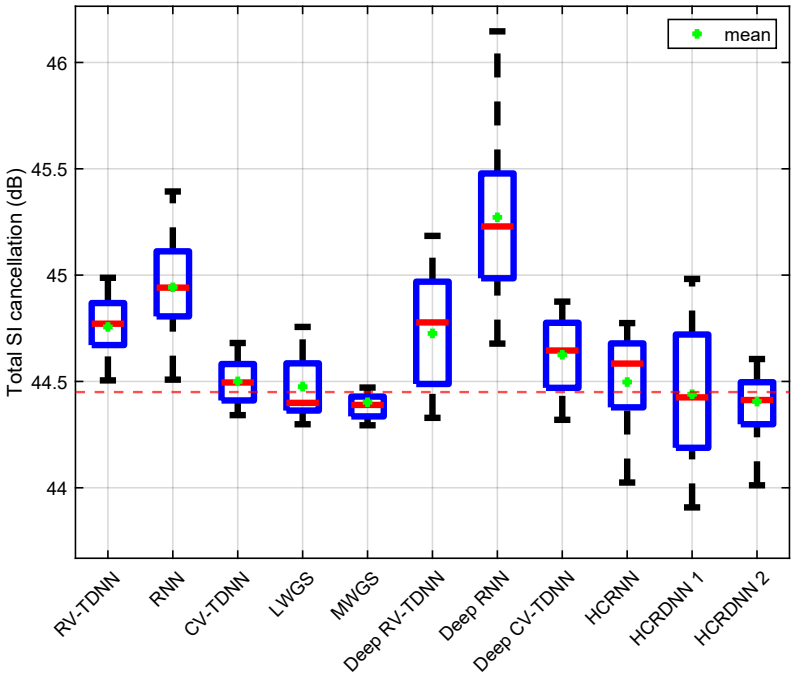


Fig. 3.8: Total SI cancellation of the proposed and the state-of-the-art NN-based cancelers.

in both training and testing phases. Additionally, it can be seen from the figures that the proposed NNs converge to similar, low MSE values in both training and testing phases, which indicates the goodness of the solution provided by the proposed NN architectures.

3.7.3 Achieved SI Cancellation

The boxplots quantifying the total SI cancellation achieved by the proposed and the state-of-the-art NN-based canceler is depicted in Fig. 3.8. Here, I calculate the mean cancellation provided by the NN-based cancelers over the employed 15 network initializations. As seen from the figure, all NNs attain a comparable cancellation performance to the polynomial canceler at $P = 5$ (red dashed-line in Fig. 3.8). Specifically, the mean cancellation achieved by all NN-based cancelers varies from 44.40 to 45.27 dB, which is very close to the polynomial target cancellation. It is

worth noting that I configure the NN settings to achieve a comparable cancellation performance to the polynomial canceler as it is not easy to adjust the setting to exactly achieve the target cancellation.

3.7.4 PSD Performance

In this subsection, I evaluate the PSD of the SI signal before applying any SI cancellation techniques (blue curve in Fig. 3.9). Further, the PSD of the residual SI signal after the linear cancellation process is assessed (red curve). Similarly, I depict the PSD of the SI signal after performing non-linear cancellation using the polynomial and NN-based cancelers. Finally, I show the PSD of the receiver background noise when there is no transmission, i.e., the receiver's noise floor (black curve). As can be inferred from Fig. 3.9, the linear canceler provides 37.90 dB SI cancellation by bringing the SI signal's power down from -42.74 dBm to -80.60 dBm. Moreover, the polynomial-based canceler attains an additional 6.6 dB cancellation by bringing the residual SI signal's power down from -80.60 dBm to -87.19 dBm, making it very close to the receiver noise floor (approximately 3 dB above the receiver's noise floor). A similar task is performed by the proposed and the state-of-the-art NN-based cancelers, in which they cancel the SI signal after the linear canceler by 6.55-7.40 dB, making it very akin to the receiver's background noise level as illustrated in the inset graph of Fig. 3.9.

3.7.5 Computational Complexity and Memory Requirements

In this subsection, I assess the computational complexity of the proposed and the state-of-the-art NN-based cancelers in terms of the number of FLOPs and calculate the complexity reduction provided by each canceler compared to the polynomial canceler. Similarly, I quantify the memory requirements of each NN-based canceler in terms

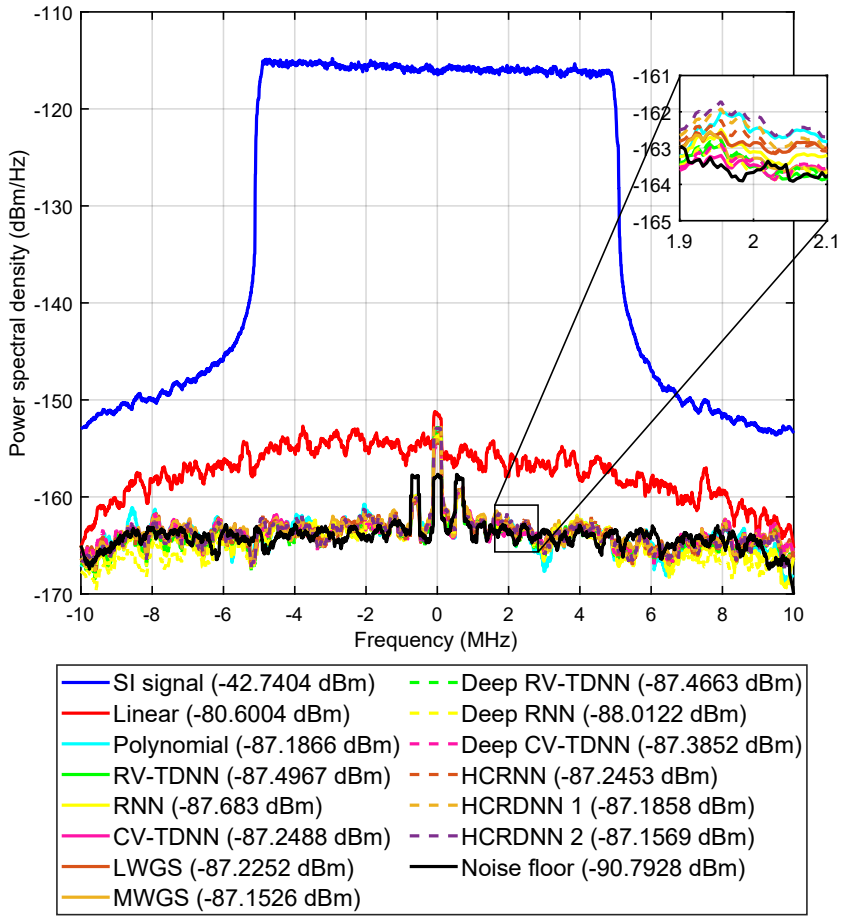


Fig. 3.9: PSD of the modeled SI signal using the proposed and the state-of-the-art NN-based cancelers.

of the number of parameters and calculate the amount of reduction compared to the baseline canceler. The results of the comparison are graphically shown in Fig. 3.10 and numerically summarized in Table 3.11.

3.7.5.1 Computational Complexity

Table 3.11 illustrates the reduction in the number of FLOPs provided by the proposed HCRNN and HCRDNN-based cancelers compared to the polynomial and the state-of-the-art NN-based cancelers. As seen from Table 3.11, the proposed NN-based cancelers achieve a superior enhancement in the computational complexity compared

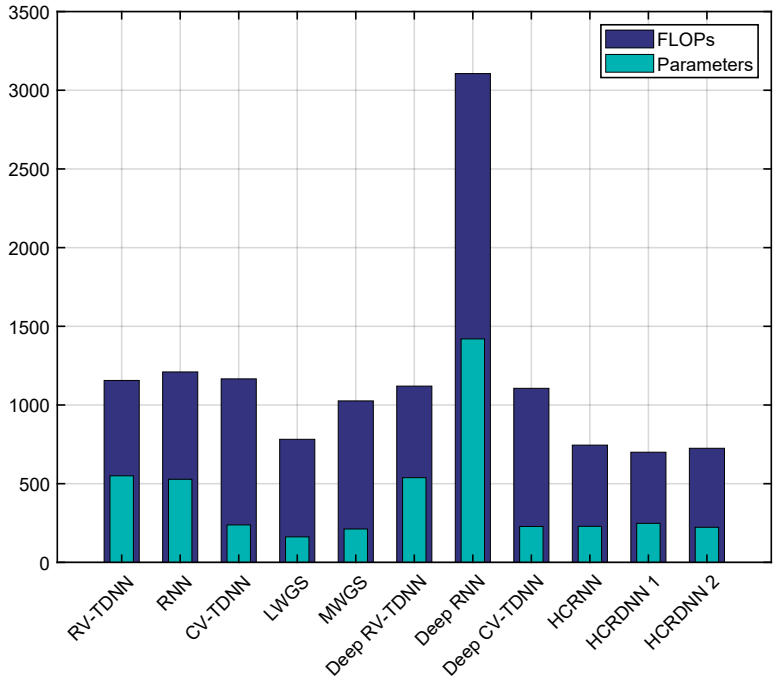


Fig. 3.10: Computational complexity and memory requirements of the proposed and the state-of-the-art NN-based cancelers.

to the polynomial-based canceler. For instance, the HCRNN provides a 52.18% reduction in the number of FLOPs while maintaining the same cancellation performance of the polynomial canceler. In addition, the HCRDNN 1 and HCRDNN 2 achieve 55.07% and 53.47% reduction compared to the polynomial canceler, respectively. However, the shallow and deep RV-TDNN, RNN, and CV-TDNN barely attain one half of the complexity reduction provided by the proposed NN architectures.

On the other hand, the proposed HCRNN, HCRDNN 1, and HCRDNN 2 architectures outperform the state-of-the-art MWGS by providing 18%, 20.92%, and 19.32% more reduction in FLOPs, respectively. Further, they attain 2.4%, 5.3%, and 3.7% more saving in FLOPs compared to the LWGS-based canceler, respectively. The above-mentioned results substantiate the proposed cancelers' superiority in modeling the SI with low computational complexity compared to the polynomial and the state-of-the-art NN-based cancelers.

TABLE 3.11: Complexity of the proposed and the state-of-the-art NN-based cancelers compared to the polynomial canceler with $P = 5$.

Network	**SI Canc. (dB)**	**Complexity**		**Complexity increase/reduction**	
		# Par.	**# FLOPs**	**# Par.**	**# FLOPs**
Polynomial ($P=5$)	44.45	312	1558	-	-
RV-TDNN	44.76	550	1156	+76.28%	-25.80%
RNN	44.94	528	1210	+69.23%	-22.34%
CV-TDNN	44.50	238	1166	-23.72%	-25.16%
LWGS	44.48	162	782	-48.08%	-49.81%
MWGS	44.40	212	1026	-32.05%	-34.15%
Deep RV-TDNN	44.73	538	1120	+72.44%	-28.11%
Deep RNN	45.27	1420	3106	+355.13%	+99.36%
Deep CV-TDNN	44.63	228	1106	-26.92%	-29.01%
HCRNN	44.50	229	745	-26.60%	-52.18%
HCRDNN 1	44.44	248	700	-20.51%	-55.07%
HCRDNN 2	44.41	223	725	-28.53%	-53.47%

3.7.5.2 Memory Requirements

Similarly, the reduction in the number of parameters of the proposed HCRNN and HCRDNN-based cancelers compared to the polynomial and the state-of-the-art NN-based cancelers is illustrated in Table 3.11. As can be observed, the shallow and deep structures of the RV-TDNN and RNN significantly increase the required parameters compared to the polynomial canceler. The former comes from dealing with the real and imaginary parts separately, while the latter results from employing a large number of feedback connections to efficiently detect the data sequences. On the other hand, the proposed HCRNN, HCRDNN 1, and HCRDNN 2 reduce the number of parameters by 26.60%, 20.51%, and 28.53% compared to the polynomial canceler, respectively. This reduction is due to the application of a convolutional layer before the recurrent layer in the proposed NN architectures, which mitigates the huge memory requirements of the recurrent layer. It can also be seen from Table 3.11 that the reduction in parameters provided by the proposed NNs is comparable to that attained

by the shallow and deep CV-TDNN. Further, the LWGS and MWGS provide the most reduction in the number of parameters.

However, the proposed NNs outperform the LWGS and MWGS in the number of FLOPs. More importantly, it can be inferred from Table 3.11 that among all NN architectures, the proposed HCRDNN 1 attains the best reduction in FLOPs. Furthermore, the proposed HCRNN achieves the best compromise among the cancellation performance, number of FLOPs, and parameters. As such, the proposed NN-based cancelers offer high design flexibility for hardware implementation, depending on the system demands.

3.8 Future Research Directions

In this work, the proposed NN architectures have been verified using a dataset that is captured by a single-input single-output FD test-bed. The joint design of multiple-input multiple-output (MIMO) and FD should be considered for B5G wireless networks [37]-[39], and the prospects of the proposed NNs could be generalized and verified using datasets captured by massive MIMO FD test-beds. In this context, the spatial correlation between the transmit and receive antennas can be exploited to design low-complexity NN-based SI cancelers [40]. Nevertheless, several challenges need to be solved prior to the deployment and design of such a system, which include but are not limited to the self- and cross-talks among the transceiver's antennas [16], [40]. Such challenges deserve a full investigation and will be studied in future work.

On the other hand, applying different machine learning techniques, such as support vector machines (SVMs) [41], for SI cancellation in FD systems is identified as another future direction of investigation. Several challenges related to the high computational complexity of the SVM models can be considered in the future.

3.9 Conclusion

In this chapter, two hybrid-layers NN architectures, referred to as hybrid-convolutional recurrent NN (HCRNN) and hybrid-convolutional recurrent dense NN (HCRDNN), have been proposed for the first time to model the FD system SI with reduced computational complexity. The former exploits the weight sharing characteristics and dimensionality reduction potential of the convolutional layer to extract the memory effect and non-linearity incorporated into the input signal using a reduced network scale. Moreover, it employs the high modeling capabilities of the recurrent layer to help learn the temporal behavior of the input signal. The latter exploits an additional dense layer to build a deeper NN model with low complexity. The complexity analysis of the proposed NN architectures has been conducted, and the optimum settings for their training have been selected. Our findings demonstrate that the proposed HCRNN and HCRDNN-based cancelers attain a reduction in the computational complexity with 52% and 55% over the polynomial-based canceler, respectively, while maintaining the same cancellation performance. In addition, the proposed HCRNN and HCRDNN offer astounding complexity reduction over the shallow and deep NN-based cancelers in the literature.

Appendix: Representation of Non-linear Systems using Polynomial based Models

In the following, I review the common methods for representing non-linear systems using polynomial-based models (e.g., Hammerstein and parallel-Hammerstein), which are considered the cornerstone of approximating the SI in FD transceivers. Herein, it is assumed that all signals are of a narrow-band nature [34]. In addition, only the odd

order non-linearities are considered since the even order non-linearities lay outside the passband [34] and are typically filtered by the BPFs in the FD transceiver [32].

A. Hammerstein Model

The Hammerstein model is one of the widely-known models for approximating the non-linear behavior with M-tap memory. In the Hammerstein model, a static non-linearity $N(.)$ is employed in series with a linear filter $L(.)$ in order to model the non-linearity with memory as follows [34]:

$$y(n) = L\left[N\left[x\left(n\right)\right]\right] = \sum_{m=0}^{M} h_m \sum_{\substack{p=1,\\ p\,odd}}^{P} a_p\, x(n-m)\,|\,x(n-m)|^{p-1}, \tag{3.46}$$

where P indicates the order of non-linearity, whereas h_m and a_p represent the coefficients of the linear filter $L(.)$ and the non-linearity $N(.)$, respectively.

B. Parallel-Hammerstein (PH) Model

An extended version of the Hammerstein model is the PH model, which is constructed by combining the outputs of several Hammerstein models and can be identified as [34]:

$$y(n) = \sum_{\substack{p=1,\\ p\,odd}}^{P} \sum_{m=0}^{M} h_{m,p}\, x(n-m)\,|x(n-m)|^{p-1}\,, \tag{3.47}$$

where $h_{m,p}$ represent the coefficients of the linear filter corresponding to that order of non-linearity.

The PH model (3.47) can be rewritten as [16] as follows:

$$y(n) = \sum_{\substack{p=1,\\ p\,odd}}^{P} \sum_{m=0}^{M} h_{m,p}\, x(n-m)^{\frac{p+1}{2}}\, x^*(n-m)^{\frac{p-1}{2}}, \tag{3.48}$$

where $(.)^*$ indicates the complex conjugate operator. The PH model (3.48) provides a more general polynomial representation for approximating the non-linearity with memory and is considered the cornerstone of modeling the SI in FD transceivers.

References

[1] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, Mar. 2020.

[2] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May 2020.

[3] H. V. Nguyen, V. Nguyen, O. A. Dobre, S. K. Sharma, S. Chatzinotas, B. Ottersten, and O. Shin, "On the spectral and energy efficiencies of full-duplex cell-free massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1698–1718, Jun. 2020.

[4] 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, Sept. 2014.

[5] D. Kim, H. Lee, and D. Hong, "A survey of in-band full-duplex transmission: From the perspective of PHY and MAC layers," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2017–2046, Feb. 2015.

[6] Z. Zhang, K. Long, A. V. Vasilakos, and L. Hanzo, "Full-duplex wireless communications: Challenges, solutions, and future research directions," *Proc. IEEE*, vol. 104, no. 7, pp. 1369–1409, Jul. 2016.

[7] 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.

[8] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

[9] E. Everett, A. Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 680–694, Feb. 2014.

[10] 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. Comput.*, Nov. 2013, pp. 992–997.

[11] M. Duarte and A. Sabharwal, "Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results," in *Proc. 44th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2010, pp. 1558–1562.

[12] Y.-S. Choi and H. Shirani-Mehr, "Simultaneous transmission and reception: Algorithm, design and system level performance," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 12, pp. 5992–6010, Dec. 2013.

[13] B. Debaillie et al., "Analog/RF solutions enabling compact full-duplex radios," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1662–1673, Sep. 2014.

[14] L. Laughlin, M. A. Beach, K. A. Morris, and J. L. Haine, "Optimum single antenna full duplex using hybrid junctions," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1653–1661, Sep. 2014.

[15] E. Ahmed and A. M. Eltawil, "All-digital self-interference cancellation technique for full-duplex systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3519–3532, Jul. 2015.

[16] D. Korpi, L. Anttila, and M. Valkama, "Nonlinear self-interference cancellation in MIMO full-duplex transceivers under crosstalk," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, pp. 1–15, Dec. 2017.

[17] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[18] M. Gao, Y. Li, O. A. Dobre, and N. Al-Dhahir, "Joint blind identification of the number of transmit antennas and MIMO schemes using Gerschgorin radii and FNN," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 373–387, Jan. 2019.

[19] H. Ye, G. Y. Li, and B.-H. F. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018.

[20] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 852–855, Oct. 2018.

[21] X. Cheng, D. Liu, C. Wang, S. Yan, and Z. Zhu, "Deep learning-based channel estimation and equalization scheme for fbmc/oqam systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 881–884, Jun. 2019.

[22] T. Gruber, S. Cammerer, J. Hoydis, and S. T. Brink, "On deep learning-based channel decoding," in *Proc. IEEE 51st Annu. Conf. Inf. Sci. Syst. (CISS)*, Baltimore, MD, USA, Mar. 2017, pp. 1–6.

[23] T. Liu, S. Boumaiza, and F. M. Ghannouchi, "Dynamic behavioral modeling of 3G power amplifiers using real-valued time delay neural networks," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 3, pp. 1025– 1033, Mar. 2004.

[24] F. Mkadem and S. Boumaiza, "Physically inspired neural network model for RF power amplifier behavioral modeling and digital predistortion," *IEEE Trans. Microw. Theory Technol.*, vol. 59, no. 4, pp. 913–923, Apr. 2011.

[25] 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.

[26] C. Tarver, L. Jiang, A. Sefidi, and J. Cavallaro, "Neural network DPD via backpropagation through a neural network model of the PA," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 358–362.

[27] 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.

[28] C. Häger and H. D. Pfister, "Nonlinear interference mitigation via deep neural networks," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, San Diego, CA, USA, Mar. 2018, pp. 1–3.

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

[30] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Advanced machine learning techniques for self-interference cancellation in full-duplex radios," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 1149–1153.

[31] 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. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 589–593.

[32] 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.

[33] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Low complexity neural network structures for self-interference cancellation in full-duplex radio," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 181-185, Jan. 2021.

[34] M. Isaksson, D. Wisell, and D. Ronnow, "A comparative analysis of behavioral models for RF power amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 1, pp. 348–359, Jan. 2006.

[35] R. Zayani, R. Bouallegue, and D. Roviras, "Adaptive predistortions based on neural networks associated with levenberg-marquardt algorithm for satellite down links," *EURASIP J. Wireless Commun. Net.,* vol. 2008, no. 1, p. 2, Jul. 2008.

[36] J. X. Peng, K. Li, and D. S. Huang, "A hybrid forward algorithm for RBF neural network construction," *IEEE Trans. Neural Netw.*, vol. 17, no. 6, pp. 1439–1451, Nov. 2006.

[37] X. Xia, K. Xu, D. Zhang, Y. Xu, and Y. Wang, "Beam-domain full-duplex massive MIMO: Realizing co-time co-frequency uplink and downlink transmission in the cellular system," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 8845–8862, Oct. 2017.

[38] K. Xu, Z. Shen, Y. Wang, X. Xia, and D. Zhang, "Hybrid time-switching and power splitting SWIPT for full-duplex massive MIMO systems: A beam-domain approach," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7257–7274, Aug. 2018.

[39] X. Xia, K. Xu, Y. Wang, and Y. Xu, "A 5G-enabling technology: Benefits, feasibility, and limitations of in-band full-duplex mmimo," *IEEE Veh. Technol. Mag.*, vol. 13, no. 3, pp. 81–90, Sep. 2018.

[40] Y. Chen et. al, "MIMO full duplex radios with deep learning," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2020, pp. 1–6.

[41] C. Auer, K. Kostoglou, T. Paireder, O. Ploder, and M. Huemer, "Support vector machines for self-interference cancellation in mobile communication transceivers," in *Proc. IEEE Veh. Technol. Conf. (VTC-Spring)*, 2020, pp. 1–6.

Chapter 4

Full-Duplex Self-Interference Cancellation Using Dual-Neurons Neural Networks

4.1 Abstract

In-band full-duplex (FD) technology has protruded as one of the most promising solutions for spectrum scarcity, by allowing users to transmit and receive simultaneously at the same center frequency. However, the FD systems suffer from a severe self-interference (SI) caused by the coupling between the transmit and receive antennas. Thereby, the potential of FD systems can not be attained without employing robust SI cancellation techniques. Traditionally, the SI is modeled using the polynomial-based cancelers, which are computationally expensive. Consequently, neural networks (NNs) have been recently introduced to model the SI with lower computational complexity. In this chapter, a novel NN structure referred to as the dual neurons-ℓ hidden layers NN (DN-ℓHLNN) is proposed. The DN-ℓHLNN exploits two neurons in the first

hidden layer to recognize the memory effect of the input and output samples with reduced complexity. The numerical simulations show that the DN-ℓHLNN-based canceler significantly reduces the computational complexity and memory requirements compared to the polynomial- and the existing NN-based cancelers while maintaining a similar non-linear cancellation performance.

4.2 Introduction

Self-interference (SI) represents a major barrier toward the practical deployment of in-band full-duplex (FD) systems [1]. Therefore, to be viable, each FD node should incorporate effective SI cancellation techniques in order to maintain the SI signal below the receiver noise floor and to avoid the saturation of the analog parts of the receiver, such as the low-noise amplifier (LNA) and the analog-to-digital converter (ADC) [1]. The cancellation of the SI could be implemented in analog and/or digital domains. Typically, analog domain cancellation is insufficient for SI suppression. As such, digital domain cancellation stages are required to suppress the remaining SI. Since the source of the SI is known to the receiver, theoretically it would be a simple task to entirely cancel the SI in the digital domain. However, this is far from practice due to the non-idealities incorporated into the input signal by various transceiver components and by the SI channel. The non-linearities of the FD transceiver are caused by the power amplifier (PA), IQ mixer, digital-to-analog converter (DAC), ADC, and LNA [2], [3].

Conventionally, the non-linearities of the FD system are approximated using the polynomial model, which can be accurate for representing the SI. However, it requires high computational complexity as the number of basis functions and model parameters significantly increases with the non-linearity order [3]. As such, neural networks (NNs)

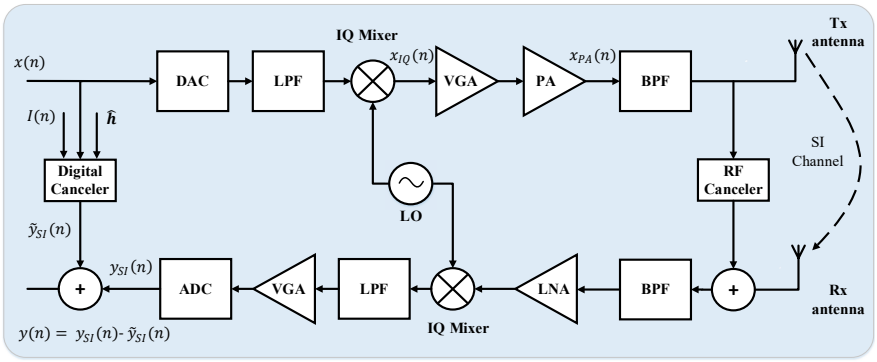
have emerged recently to model the SI with lower computational complexity.

In [2], the non-linear SI signal is modeled using the conventional real-valued feed-forward NN (RV-FFNN). Moreover, in [4], advanced NNs such as recurrent NN (RNN) and complex-valued FFNN (CV-FFNN) are introduced to estimate the SI signal's non-linear part; it is shown that the CV-FFNN is more suitable than the RV-FFNN and RNN for SI cancellation. In [5], the authors propose two CV NN structures termed as the ladder-wise grid structure (LWGS) and moving-window grid structure (MWGS) to model the non-linear part of the SI signal with lower computational complexity than the CV-FFNN introduced in [4]. In addition, in [6], two hybrid-layers NN structures, namely the hybrid-convolutional recurrent NN (HCRNN) and hybrid-convolutional recurrent dense NN (HCRDNN), have been recently introduced to modeling the non-linear SI with lower computational complexity than the CV-FFNN [4], LWGS [5], and MWGS [5]. On the other hand, in [7] and [8], an RV-FFNN known as a two-hidden layer NN (2HLNN) is used to model the non-linearity of memory-based systems such as the Doherty RF PA. In this case, the 2HLNN significantly outperforms the time-delay NN (TDNN)[1] introduced in [9].

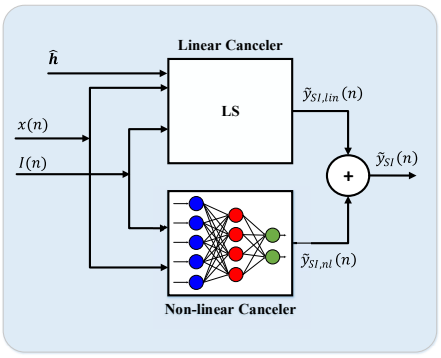
Motivation: The existing NN-based cancelers [2], [4]-[6] provide a good reduction in the complexity when compared to the traditional polynomial canceler; however, further reduction in complexity is mandated for building energy-efficient digital cancelers, which can be feasible for hardware implementation in wireless devices. Thus, designing novel NN-based cancelers with lower computation and memory requirements than the existing NN-based cancelers, and having the scope of practical applicability, is the main motivation of this chapter.

Contributions: The major contributions of this chapter are summarized as follows.

[1]It is noted that the FFNN implemented in [2] has a similar structure to the TDNN introduced in [9] since both employ the buffered samples of the input signal at the input layer. Henceforward, I will use the term TDNN instead of FFNN for the sake of accurate terminology.



(a) Detailed system model.



(b) Digital canceler.

Fig. 4.1: Full-duplex transceiver system model.

Firstly, I investigate the 2HLNN for SI cancellation in FD systems, but by applying the CV framework instead of the traditional RV employed in [7], due to the potential complexity reduction [4]. Secondly, I propose a novel CV NN structure, known as the dual neurons-ℓ hidden layers NN (DN-ℓHLNN), to model the SI with lower complexity and memory requirements than the polynomial- and the existing NN-based cancelers.

4.3 System Model

The FD transceiver system emphasizing the SI signal path is shown in Figs. 4.1(a) and (b). In this work, the typical polynomial model is employed as a benchmark to approximate the SI signal. To derive this model, it is considered that the PA and IQ mixer are the main sources for non-linearities in the FD transceiver [10].

Polynomial-based SI Signal Representation: The digital baseband signal $x(n)$, shown in Fig. 4.1(a), is converted to analog using the DAC, filtered by the low-pass filter (LPF), and then up-converted using the IQ mixer, which introduces IQ imbalance to the input signal. The digital equivalent of the output signal after the IQ mixer can be expressed as [3]

$$x_{IQ}(n) = \frac{1}{2}(1 + \psi e^{j\theta})\ x(n) + \frac{1}{2}(1 - \psi e^{j\theta})\ x^*(n), \tag{4.1}$$

with θ and ψ as the phase and gain imbalance components of the transmitter, respectively. The mixed-signal amplitude is adjusted using the variable gain amplifier (VGA) before it is amplified by the PA, which exhibits an additional non-linearity to the input signal. This non-linearity can be modeled using the parallel Hammerstein model as [3]

$$x_{PA}(n) = \sum_{\substack{p=1,\\ p\,odd}}^{P} \sum_{m=0}^{M} h_{m,p}\, x_{IQ}(n-m)^{\frac{p+1}{2}} x_{IQ}^*(n-m)^{\frac{p-1}{2}}, \tag{4.2}$$

where $h_{m,p}$ indicates the parallel Hammerstein model's impulse response, while M and P are the memory depth and non-linearity order of the PA, respectively.

The output signal of the PA is filtered using the band-pass filter (BPF) before passing through the SI channel. It is assumed, for simplicity of exposition, that no signal of interest (SoI) exists from a remote FD node and no thermal noise [2], [10]. Accordingly, the residual SI signal, after employing RF cancellation using antenna separation and/or using adaptive techniques such as [1], is filtered, amplified, down converted, and digitized using the BPF, LNA, IQ mixer, and ADC, respectively, and is expressed as [2]

$$y_{SI}(n) = \sum_{\substack{p=1,\\ p\,odd}}^{P} \sum_{q=0}^{p} \sum_{m=0}^{M_i-1} h_{m,q,p}\, x(n-m)^q x^*(n-m)^{p-q}, \tag{4.3}$$

where $h_{m,q,p}$ denotes the impulse response of a channel that includes the overall effect of the SI channel, PA, and IQ mixer, whereas M_i represents the memory effect incorporated into the input signal by the SI channel and PA.

Digital Domain Cancellation: In the digital domain, the main function of the can-

celer is to provide an accurate estimate of the SI (4.3), which is denoted by $\tilde{y}_{_{SI}}(n)$ as shown in Fig. 4.1(a). This is subtracted from the SI signal, and the residual SI after the digital SI cancellation is given by $y(n) = y_{_{SI}}(n) - \tilde{y}_{_{SI}}(n)$. The SI cancellation over a K-length window is calculated in dB as $\xi_{dB} = 10\log_{10}\left(\sum_{n=0}^{K-1}|y_{_{SI}}(n)|^2 \,/\, \sum_{n=0}^{K-1}|y(n)|^2\right)$.

Linear and Non-Linear Cancellations: In this work, the digital SI signal is estimated by approximating its linear and non-linear parts using the linear and non-linear cancelers, respectively, as shown in Fig. 4.1(b). The linear component of the SI can be approximated by letting $p = q = 1$ in (4.3) as

$$\tilde{y}_{_{SI,lin}}(n) = \sum_{m=0}^{M_i-1} h_{m,1,1}\, x(n-m), \tag{4.4}$$

where $h_{m,1,1}$ is estimated using the least-squares (LS) channel estimation approach as follows [6]:

$$\hat{\boldsymbol{h}} = \left(\boldsymbol{X}^{\boldsymbol{H}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{H}}\boldsymbol{y}_{_{SI}}, \tag{4.5}$$

with $(.)^H$ as the conjugate transpose, whereas $\boldsymbol{X} \in \mathbb{C}^{N\times M_i}$ and $\boldsymbol{y}_{_{SI}} \in \mathbb{C}^{N\times 1}$ are the input and output training data, respectively, with N as the length of training data. The non-linear component of the SI signal is estimated using the NN, and its value depends on the NN model, as will be shown in the next section. The approximated SI is then reconstructed by summing the linear and non-linear parts as $\tilde{y}_{_{SI}}(n) = \tilde{y}_{_{SI,lin}}(n) + \tilde{y}_{_{SI,nl}}(n)$.

4.4 Proposed NN-Based Canceler

4.4.1 NN Structure

The main target of the NN-based canceler is to detect the temporal behavior that exists in the SI signal. This motivates the use of NN structures that accommodate the SI signal's memory effect, such as RNN and TDNN. However, the training complexity of the RNN is considered impractically high [8]. Therefore, the TDNN is preferred to model the SI versus the RNN [4]. Another NN that could accommodate the SI signal's temporal behavior is the 2HLNN [7], as it is shown to outperform the TDNN in modeling the PA's non-linearity. However, the 2HLNN can come at a cost in computational complexity since the delayed versions of both input and output signals are fully connected to the first hidden layer neurons [7].

To relax the computational complexity of the 2HLNN, the DN-ℓHLNN structure, shown in Fig. 4.2, is proposed in this chapter. In the DN-ℓHLNN, only two neurons are utilized in the first hidden layer instead of an arbitrary number of neurons used in the case of 2HLNN. Moreover, in the DN-ℓHLNN, the input units are not fully connected to the first hidden layer; the input samples are connected to the first neuron only, while the output samples are connected to the second neuron. As such, the non-linear SI signal, modeled by the proposed DN-ℓHLNN structure as shown in Figs. 4.1 and 4.2, can be given as:

$$\tilde{y}_{SI,nl}(n) = g\left\{x\left(n\right), I\left(n\right), O\left(n\right)\right\}, \tag{4.6}$$

where $g\left\{.\right\}$ is the DN-ℓHLNN mapping function, while $I\left(n\right)$ can be expressed as $I\left(n\right) = \left\{x\left(n-1\right), x\left(n-2\right), ..., x\left(n-M_i+1\right)\right\}$ and $O\left(n\right)$ can be written as $O\left(n\right) = \left\{z\left(n-1\right), z\left(n-2\right), ..., z\left(n-M_o\right)\right\}$, with $z\left(n\right) = y_{SI}(n) - \tilde{y}_{SI,lin}(n)$ as the feedback output signal after applying the linear cancellation. Finally, M_i and M_o represent

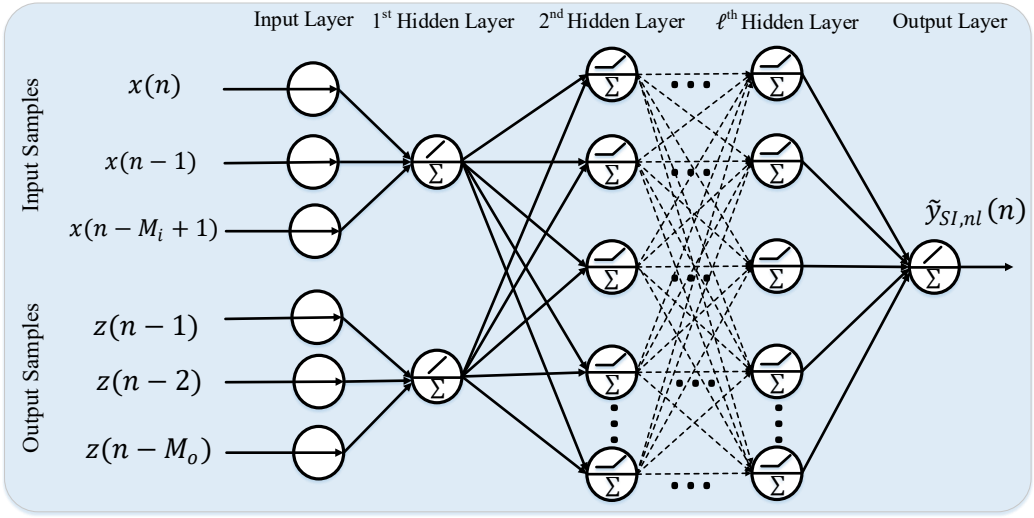


Fig. 4.2: Proposed DN-ℓHLNN structure.

the memory effect attributed to the input and output signals, respectively. Including M_o samples of the received data at the input of the DN-ℓHLNN helps to take into account the composite effect of the transceiver's non-linearities and the over-the-air SI propagation delay spread.

The main idea behind this configuration is to allow the two neurons employed in the first hidden layer to recognize the memory effect of the input and output signals separately while reducing the required number of network parameters (e.g., weights and biases). To detect the above-mentioned memory effects, the DN-ℓHLNN employs linear activation functions in the first hidden layer. On the other hand, the second, third, and ℓ^{th} hidden layers' neurons of the DN-ℓHLNN are utilized to approximate the non-linearities induced by the various components of the FD transceiver. To model these non-linearities, the DN-ℓHLNN uses non-linear activation functions in the aforementioned hidden layers, such as the complex ReLU ($\mathbb{C}$ReLU) activation function [4], [5].

4.4.2 Computational Complexity

In this subsection, I derive the computational complexity of the DN-ℓHLNN for $\ell \geq 2$. This is evaluated in terms of the total number of floating-point operations (FLOPs) that are required in the NN's inference phase. In this stage, I evaluate the RV multiplications (MULs) and additions (ADDs) utilized at the DN-ℓHLNN's neurons to calculate the weighted sum of its inputs, add the biases, and apply the activation function. Thus, the overall complexity of the DN-ℓHLNN is given as

$$\mathcal{O}_{DN-\ell HLNN} = \mathcal{O}_{w,b} + \mathcal{O}_{a}, \tag{4.7}$$

where $\mathcal{O}_{w,b} = \gamma_{w,b}^{\Re_m} + \gamma_{w,b}^{\Re_a}$ and $\mathcal{O}_{a} = \gamma_{a}^{\Re_m} + \gamma_{a}^{\Re_a}$, with $\gamma_{w,b}^{\Re_m}$ and $\gamma_{w,b}^{\Re_a}$ representing the overall RV MULs and ADDs, respectively, which are used for evaluating the weighted sum of the inputs and adding the bias terms in all neurons of the DN-ℓHLNN. Likewise, $\gamma_{a}^{\Re_m}$ and $\gamma_{a}^{\Re_a}$ denote the overall RV MULs and ADDs needed to apply the activation functions in all neurons of DN-ℓHLNN.

Since the DN-ℓHLNN deals with CV inputs and CV network parameters, all the network operations (e.g., MULs and ADDs) will be CV. I consider that each CV ADD requires two RV ADDs, and each CV MUL costs five RV ADDs and three RV MULs [10]. Accordingly, it can be easily deduced that $\gamma_{w,b}^{\Re_m}$ and $\gamma_{w,b}^{\Re_a}$ can be obtained as

$$\gamma_{w,b}^{\Re_m} = 3\gamma_{w,b}^{\mathbb{C}_m}, \quad \gamma_{w,b}^{\Re_a} = 5\gamma_{w,b}^{\mathbb{C}_m} + 2\gamma_{w,b}^{\mathbb{C}_a}, \tag{4.8}$$

where $\gamma_{w,b}^{\mathbb{C}_m}$ and $\gamma_{w,b}^{\mathbb{C}_a}$ represent the CV counterparts of $\gamma_{w,b}^{\Re_m}$ and $\gamma_{w,b}^{\Re_a}$, respectively. At the j^{th} hidden layer of the DN-ℓHLNN, $\gamma_{w,b}^{\mathbb{C}_m,j}$ and $\gamma_{w,b}^{\mathbb{C}_a,j}$ can be calculated as

$$\gamma_{w,b}^{\mathbb{C}_m,j} = \gamma_{w,b}^{\mathbb{C}_a,j} = n_{j-1}n_j, \quad 1 < j \leq \ell, \tag{4.9}$$

where n_j denotes the number of neurons in the j^{th} hidden layer of the DN-ℓHLNN. It is worth mentioning that (4.9) is applied to all the hidden layers of the DN-ℓHLNN except the first hidden layer since it is not fully connected to all the input layer's neurons. The computation of the CV operations of the first hidden layer is performed as follows. For the first (second) neuron with M_i (M_o) inputs connected to it, the number of CV MULs and ADDs is M_i (M_o), respectively, to calculate the weighted sum of its inputs and add the bias term. Thus, $\gamma_{w,b}^{\mathbb{C}_m,1} = \gamma_{w,b}^{\mathbb{C}_a,1} = M_i + M_o$.

Furthermore, since the DN-ℓHLNN employs only one neuron at the output layer, the number of CV MULs and ADDs associated with the output layer is equal to the number of neurons in the last hidden layer (i.e., n_ℓ). Thus, $\gamma_{w,b}^{\mathbb{C}_m}$ and $\gamma_{w,b}^{\mathbb{C}_a}$ can be expressed as

$$\gamma_{w,b}^{\mathbb{C}_m} = \gamma_{w,b}^{\mathbb{C}_a} = M_i + M_o + \sum_{j=2}^{\ell} n_{j-1} n_j + n_\ell. \tag{4.10}$$

Finally, the number of RV MULs and ADDs required for applying the activation functions at all neurons of the DN-ℓHLNN can be expressed as follows [5]:

$$\gamma_a^{\Re_m} = \sum_{j=2}^{\ell} 2n_j, \quad \gamma_a^{\Re_a} = \sum_{j=2}^{\ell} 6n_j. \tag{4.11}$$

Using (4.8), (4.10), and (4.11), the overall complexity of the DN-ℓHLNN (including the linear canceler complexity) can be obtained according to (4.7) as follows:

$$\mathcal{O} = 10 \left\{ M_i + M_o + \sum_{j=2}^{\ell} n_{j-1} n_j + n_\ell \right\} + \sum_{j=2}^{\ell} 8n_j + \mathcal{O}_{lin} + 2, \tag{4.12}$$

where $\mathcal{O}_{lin}$ represents the number of FLOPs of the linear canceler. The two FLOPs added in (4.12) indicate the number of RV ADDs needed to sum the outputs of the linear and non-linear cancelers as in [6], [10].

4.5 Results and Discussion

In this section, the performance of the proposed DN-ℓHLNN is evaluated and compared with that of the CV-TDNN [4], LWGS [5], MWGS [5], HCRNN [6], HCRDNN [6], and 2HLNN [7] using different network configurations. All NNs are employed to estimate the SI signal's non-linear part using CV inputs, except for the hybrid-layers structures, which employ RV inputs [6]. The NNs are developed in Python using TensorFlow and Keras libraries. Further, to train the NNs, in this work, I used the dataset[2] employed in [2] and [4], which contains measured data from a hardware FD test-bed that employs an orthogonal frequency division multiplexing signal with quadrature phase-shift keying modulation and 10 MHz pass-band bandwidth. The dataset consists of 20,480 observations, where 90% of the observations are utilized for training, while 10% are applied for testing.

The network configuration of the proposed DN-ℓHLNN (e.g., number of hidden layers, number of neurons) is adjusted using a trial and error mechanism, such that it achieves a comparable SI cancellation to the polynomial-based canceler at $P = 5$. In other words, in this work, the polynomial canceler is considered the baseline versus which the complexity of the proposed and the existing NN-based cancelers is compared. To that end, I tested a large number of configurations for the proposed DN-ℓHLNN using $\ell = 2$ and $\ell = 3$ and varied the number of neurons in the 2^{nd} and the 3^{rd} (if any) hidden layers such that $n_j \in \{3, 4..., 13\}$, with $j \in \{2, 3\}$. Among a large number of tested configurations, I select the shallow DN-2HLNN (2-6) and DN-2HLNN (2-7), for the case of $\ell = 2$, as they attain a comparable cancellation performance to that of the polynomial-based canceler with the lowest computational

[2]In this work, the system model is simplified for exposition; however, the employed dataset is coming from measurements, and hence, takes into consideration the effect of noise as well as all transceiver's non-linearities.

TABLE 4.1: Optimum settings of the proposed and the existing NNs.

Structure	M_i	M_o	**# Filters**	**Filter size**	**# Hidden layers neurons**	**Window size**	**Act.**	**Learning rate** ($\times 10^{-3}$)	**Batch size**	**Opti.**
Shallow CV-TDNN [4]	13	N/A	N/A	N/A	7	N/A	$\mathbb{C}$ReLU	4.5	62	Adam
Shallow 2HLNN [7]	13	12	N/A	N/A	(2-7)	N/A	$\mathbb{C}$ReLU	4.5	62	Adam
Shallow DN-2HLNN	13	12	N/A	N/A	(2-6)/(2-7)	N/A	$\mathbb{C}$ReLU	4.5	62	Adam
Deep CV-TDNN [4]	13	N/A	N/A	N/A	(4-4-4)	N/A	$\mathbb{C}$ReLU	4.5	62	Adam
Deep 3HLNN [7]	13	12	N/A	N/A	(2-4-5)	N/A	$\mathbb{C}$ReLU	4.5	62	Adam
Deep DN-3HLNN	13	12	N/A	N/A	(2-4-5)	N/A	$\mathbb{C}$ReLU	4.5	62	Adam
Shallow LWGS [5]	13	N/A	N/A	N/A	9	N/A	$\mathbb{C}$ReLU	4.5	62	Adam
Shallow MWGS [5]	13	N/A	N/A	N/A	12	5	$\mathbb{C}$ReLU	4.5	62	Adam
Shallow HCRNN [6]	13	N/A	3	12×1×1	9	N/A	ReLU	5	62	Adam
Deep HCRDNN 1 [6]	13	N/A	2	12×1×1	(7-11)	N/A	ReLU	5	158	Adam
Deep HCRDNN 2 [6]	13	N/A	3	12×1×1	(5-12)	N/A	ReLU	5	158	Adam

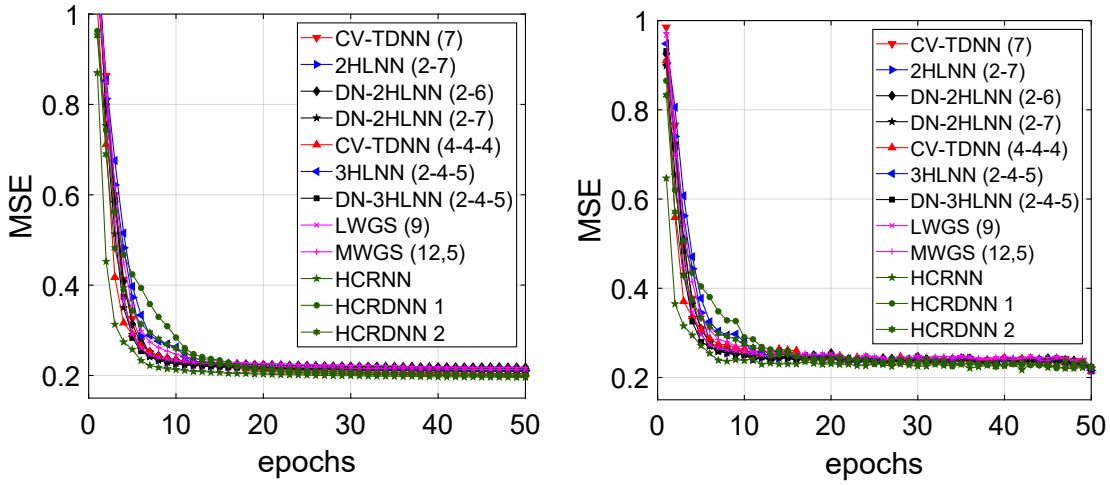
complexity. Further, for the case of $\ell = 3$, I select the deep DN-3HLNN (2-4-5).[3] Using a similar procedure, I choose the shallow 2HLNN (2-7) and the deep 3HLNN (2-4-5)[4] for $\ell = 2$ and $\ell = 3$, respectively, as they achieve the baseline cancellation with the lowest complexity.

In addition, among the traditional NNs in the literature (i.e., RV-TDNN, RNN, and CV-TDNN), the CV-TDNN is selected for comparison as it is shown to be more powerful for the SI cancellation problem than the RV-TDNN and RNN [4]. Finally, to show the advantage of the proposed NNs over the recent works in [5] and [6], I include the LWGS, MWGS, HCRNN, and HCRDNN. The network settings for training the above-mentioned NNs (e.g., learning rate, batch size, activation functions, etc.) are also optimized based on a trial and error methodology and are summarized in Table 4.1.

Figs. 4.3(a) and (b) depict the mean square error (MSE) in the training and testing phases, respectively, for the aforementioned NNs using 15 random initializations. As illustrated in Fig. 4.3, the MSE curves of all NNs are very akin to each other since

[3]Interestingly, It is noted that increasing the number of hidden layers of the proposed DN-ℓHLNN over $\ell = 3$ slightly enhances the cancellation performance at the cost of a significant increase in complexity.

[4]It is noted that the 3HLNN structure is not considered in [7], and it is investigated for the first time in this chapter.



(a) Training phase. (b) Testing phase.

Fig. 4.3: MSE performance for different NN structures.

the NNs are configured to attain a similar cancellation performance. Moreover, it can be seen from Figs. 4.3(a) and (b) that the MSE values for training and testing phases are comparable for each NN, and there is no case of overfitting for any of the NN structures.

The boxplots of the non-linear cancellation provided by each NN-based canceler over the different seed initializations are shown in Fig. 4.4. As can be seen from this figure, the mean non-linear cancellations achieved by all NNs are comparable to that of the polynomial canceler at $P = 5$ (dashed red-line in Fig. 4.4). Further, it can be observed from the boxplots in Fig. 4.4 that the non-linear cancellation of the proposed DN-ℓHLNN structures exhibits good stability over the different seed initializations.

The power spectral density (PSD) of the SI signal after applying the linear and non-linear cancellations is shown in Fig. 4.5. As seen from the figure, the LS-based linear canceler provides 37.86 dB linear cancellation, whereas the proposed DN-ℓHLNN-based non-linear cancelers (i.e., DN-2HLNN and DN-3HLNN) achieve around 6.6 dB non-linear cancellation, and thus, 44.45 dB total cancellation. This amount of total cancellation is comparable to that of the polynomial- and the existing NN-based

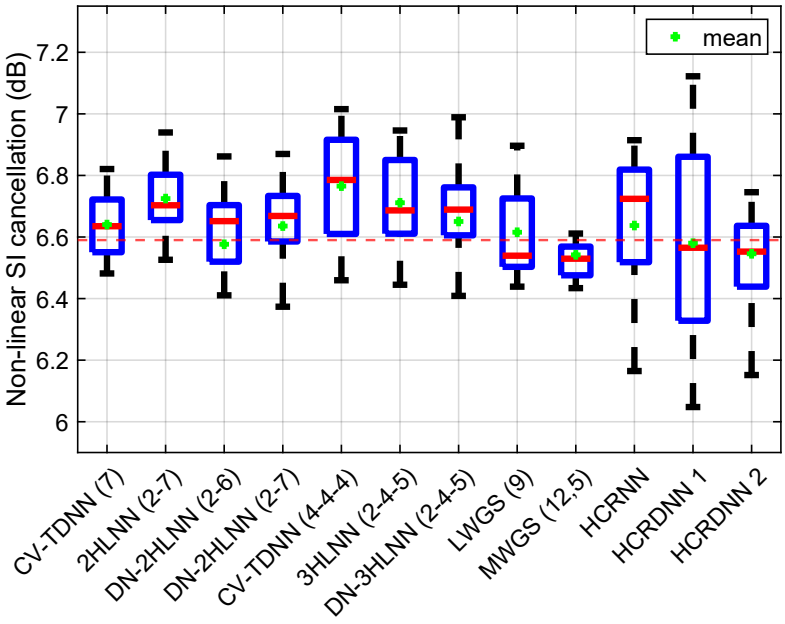


Fig. 4.4: Non-linear SI cancellation on the test data for various NN structures.

cancelers. As such, the gap to the receiver noise floor provided by the polynomial- and all NN-based cancelers are comparable subsequently, as can be seen from Fig. 4.5. Finally, it can be observed from Fig. 4.5 that the residual SI signal's spectra after applying the proposed NNs is similar to that of the receiver noise floor, which reveals the high modeling capabilities of the proposed structures to suppress the SI to the receiver's noise level.

In Table 4.2, the complexity of the proposed and the existing NNs is evaluated in terms of the number of network parameters and FLOPs and compared with the complexity of the polynomial canceler at a similar cancellation performance. As seen from the table, in the case of $\ell = 2$, the shallow DN-2HLNN (2-6) and DN-2HLNN (2-7) attain the best reduction in the complexity among all other shallow and deep NNs in the literature; they reduce the number of parameters and FLOPs by 58% and 61%; 56% and 59%, respectively, compared to the polynomial-based canceler at $P = 5$. It is worth noting that increasing the number of neurons from six to seven

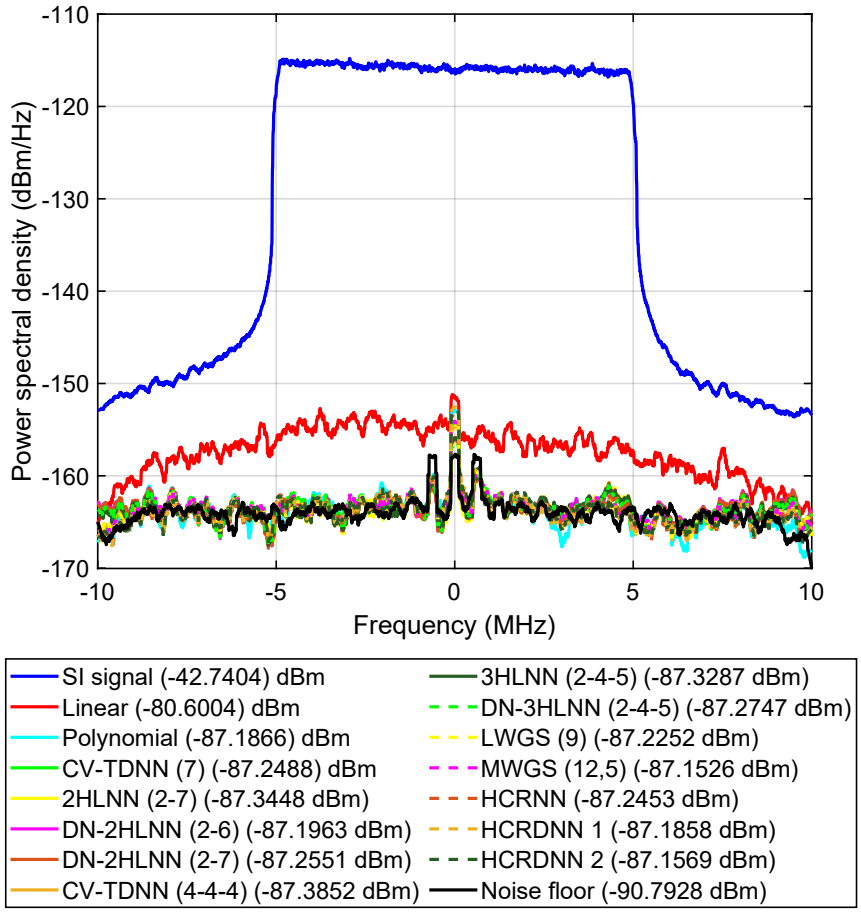


Fig. 4.5: PSD of the SI signal modeled by various NN structures.

for the DN-2HLNN structure provides an additional SI cancellation of 0.06 dB at the expense of increasing both the number of parameters and FLOPs by 2.5%. This indicates the reliability of the proposed DN-2HLNN from both SI cancellation and complexity reduction perspectives.

On the other hand, for $\ell = 3$, the deep DN-3HLNN (2-4-5) attains a significant reduction in the parameters compared to its deep (three-hidden layers) NN counterparts, i.e., CV-TDNN (4-4-4), 3HLNN (2-4-5), HCRDNN 1, and HCRDNN 2; to this end, it provides more than half of the reduction achieved by the most efficient structure from this group of NNs (i.e., 3HLNN (2-4-5)). However, in terms of FLOPs, it slightly requires more operations than the most efficient structure from the same group (i.e.,

TABLE 4.2: Complexity reduction for various NN structures.

Network	LSIC (dB)	NSIC (dB)	TSIC (dB)	GAP2RX	Complexity		Reduction	
					# Par.	# FLOPs	# Par.	# FLOPs
Polynomial ($P=5$)	37.86	6.59	44.45	3.61	312	1558	-	-
CV-TDNN (7)	37.86	6.64	44.50	3.54	238	1166	-23.72%	-25.16%
2HLNN (2-7)	37.86	6.72	44.58	3.45	188	896	-39.74%	-42.49%
DN-2HLNN (2-6)	**37.86**	**6.58**	**44.44**	**3.60**	**130**	**608**	**-58.33%**	**-60.98%**
DN-2HLNN (2-7)	**37.86**	**6.64**	**44.50**	**3.54**	**138**	**646**	**-55.77%**	**-58.54%**
CV-TDNN (4-4-4)	37.86	6.77	44.63	3.41	228	1106	-26.92%	-29.01%
3HLNN (2-4-5)	37.86	6.71	44.57	3.46	216	1032	-30.77%	-33.76%
DN-3HLNN (2-4-5)	**37.86**	**6.65**	**44.51**	**3.52**	**166**	**782**	**-46.79%**	**-49.81%**
LWGS (9)	37.86	6.62	44.48	3.57	162	782	-48.08%	-49.81%
MWGS (12,5)	37.86	6.54	44.40	3.64	212	1026	-32.05%	-34.15%
HCRNN	37.86	6.64	44.50	3.55	229	745	-26.60%	-52.18%
HCRDNN 1	37.86	6.58	44.44	3.60	248	700	-20.51%	-55.07%
HCRDNN 2	37.86	6.55	44.41	3.64	223	725	-28.53%	-53.47%

LSIC: linear SIC; NSIC: non-linear SIC; TSIC: total SIC; GAP2RX: gap to Rx noise floor.

HCRDNN 1). Thus, in sum, one can use the shallow DN-2HLNN (2-6) rather than the deep DN-3HLNN (2-4-5) as it requires lower memory and computations than all shallow and deep NN structures.

4.6 Conclusion

In this chapter, a novel NN structure referred to as the DN-ℓHLNN has been proposed for modeling the SI in an FD transceiver. The DN-ℓHLNN reduces the computational complexity versus the existing polynomial and NN-based cancelers while achieving a similar cancellation performance. The simulation results have demonstrated that, using $\ell = 2$, the DN-2HLNN-based canceler provides about 60% reduction in the number of network parameters and FLOPs over the polynomial-based canceler. Furthermore, the DN-2HLNN has attained a superior performance enhancement in parameters and FLOPs over the existing NN-based cancelers. Future works can consider testing the performance of the proposed NNs using a dataset captured by a multiple-input multiple-output FD testbed. Further, the effect of the SoI on the system performance can also be investigated in the future.

References

[1] A. T. Le *et* al., "Analog least mean square adaptive filtering for self-interference cancellation in full duplex radios," *IEEE Wireless Commun.*, vol. 28, no. 1, pp. 12–18, Feb. 2021.

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

[3] D. Korpi, L. Anttila, and M. Valkama, "Nonlinear self-interference cancellation in MIMO full-duplex transceivers under crosstalk," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, pp. 1–15, Dec. 2017.

[4] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Advanced machine learning techniques for self-interference cancellation in full-duplex radios," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 1149–1153.

[5] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Low complexity neural network structures for self-interference cancellation in full-duplex radio," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 181-185, Jan. 2021.

[6] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Hybrid-layers neural network architectures for modeling the self-interference in full-

duplex systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6291–6307, Jun. 2022.

[7] F. Mkadem *et* al., "Behavioral modeling and digital predistortion of power amplifiers with memory using two hidden layers artificial neural networks," in *Proc. IEEE MTT-S Int. Microw. Symp.*, 2010, pp. 656–659.

[8] F. Mkadem and S. Boumaiza, "Physically inspired neural network model for RF power amplifier behavioral modeling and digital predistortion," *IEEE Trans. Microw. Theory Technol.*, vol. 59, no. 4, pp. 913–923, Apr. 2011.

[9] T. Liu *et* al., "Dynamic behavioral modeling of 3G power amplifiers using real-valued time delay neural networks," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 3, pp. 1025–1033, Mar. 2004.

[10] Y. Kurzo *et* al., "Hardware implementation of neural self-interference cancellation," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 10, no. 2, pp. 204-216, Jun. 2020.

Chapter 5

Machine Learning-based Self-Interference Cancellation for Full-Duplex Radio: Approaches, Open Challenges, and Future Research Directions

5.1 Abstract

In contrast to the long-held belief that wireless systems can only work in half-duplex mode, full-duplex (FD) systems are able to concurrently transmit and receive information over the same frequency bands to theoretically enable a twofold increase in spectral efficiency. Despite their significant potential, FD systems suffer from an inherent self-interference (SI) due to a coupling of the transmit signal to its own FD receive chain. Self-interference cancellation (SIC) techniques are the key enablers

for realizing the FD operation, and they could be implemented in the propagation, analog, and/or digital domains. Particularly, digital domain cancellation is typically performed using model-driven approaches, which have proven to be insufficient to seize the growing complexity of forthcoming communication systems. For the time being, machine learning (ML) data-driven approaches have been introduced for digital SIC to overcome the complexity hurdles of traditional methods. This chapter reviews and summarizes the recent advances in applying ML to SIC in FD systems. Further, it analyzes the performance of various ML approaches using different performance metrics, such as the achieved SIC, training overhead, memory storage, and computational complexity. Finally, this chapter discusses the challenges of applying ML-based techniques to SIC, highlights their potential solutions, and provides a guide for future research directions.

5.2 Introduction

The sixth-generation (6G) wireless networks are anticipated to connect "intelligence" rather than "things" while maintaining the quality-of-service requirements of low latency, massive connectivity, and stringent energy efficiency [1]-[20]. Through several technologies, 6G visionaries expect an unprecedented provision of services to 6G users by allowing 10 times lower latency, 100 times higher connectivity, and 1000 times higher data rates compared to the fifth-generation wireless systems' users [1]-[4].

To meet the high data rate requirements of 6G networks, the in-band full-duplex (IBFD) systems have emerged as one of the potential technologies owing to their ability to serve a large number of devices concurrently on the same frequency bands [21]-[43]. Given this potential, the IBFD devices can theoretically provide a twofold increase in spectral efficiency, making them promising candidates for 6G networks.

Doubling the spectral efficiency, however, comes at the cost of having an inevitable self-interference (SI) at the receiver (Rx) chain of an FD node from its own transmitter (Tx) chain. To break through such a bottleneck, SI cancellation (SIC) has been verified as the panacea that can enable the essence of IBFD communications [21], [25], [26], [30], [34].

In the past few decades, researchers have drawn attention to canceling the SI in IBFD systems.[1] Generally, the SIC can be performed in propagation, analog, and/or digital domains. Propagation domain cancellation can be performed at the radio frequency (RF) level using antenna isolation [21], beamforming [28], polarized antennas [44], circulators [45], and/or hybrid junction networks [46]. On the other hand, analog domain cancellation can be carried out actively by generating a pre-processed copy of the SI signal, which is exploited to cancel the original SI signal at the Rx chain. Analog domain cancellations are often incapacitated to suppress the SI signal to the Rx noise floor level. As a consequence, additional focus has been directed to canceling the SI at the baseband level using digital domain cancellation [47]-[56]. At low or moderate transmit power levels, the digital domain cancellation is typically performed using linear cancelers, which reconstruct an estimated copy of the SI signal based on techniques such as least-squares (LS) channel estimation [47], [49], [53]. However, at high transmit power levels, such cancellation only becomes insufficient to entirely suppress the SI to the Rx noise floor due to the stringent non-linear behavior of FD transceiver's components, such as the power and low-noise amplifiers (PA and LNA) [47], [49], [52]. Thus, non-linear digital cancellation is applied with the linear cancellation to bring the SI to the Rx noise floor level. The non-linear SIC is conventionally

[1]Investigating the SIC for sub-band FD, where a single frequency band is partitioned into sub-bands for down-link and up-link transmissions, and both can take place simultaneously, is an important point that can be considered for future explorations. Further, investigating the SIC for wide-band FD, where the FD capability is to be achieved over wide bandwidths, can also be considered in the future.

performed using model-driven approaches, e.g., polynomial models, which are shown to fit well in practice; however, they need many trainable parameters that, in turn, translate to higher computational requirements [57].

Artificial intelligence, a wide-ranging area of computer science, has currently made a profound technological revolution in all disciplines of communications [58]-[87]. Specifically, machine and deep learning (ML and DL), aiming to extract hidden features, i.e., insights, from training data, have attained considerable success in channel coding [64], [65], channel estimation [73], [80], [85], [86], channel equalization [73], signal identification [70], [87], signal detection [63], optical fiber's signal-to-noise ratio estimation [82], [83], digital pre-distortion [88], and PA behavioral modeling [89]. In these works, the data-driven ML approaches have achieved astonishing enhancements in either performance or complexity when compared to the model-driven approaches.

Applying ML to IBFD communications has recently been regarded as one of the promising techniques that supports the horizon of 6G networks [90]-[126]. To that extent, traditional ML techniques, such as neural networks (NNs) and support vector regressors (SVRs), have been introduced for SIC in FD transceivers [90]-[108]. Further, advanced ML techniques, such as tensor completion (TC), TensorFlow graphs, and so forth, have also been investigated for learning the SI in FD transceivers [109], [110], [112]. Other ML techniques, such as Gaussian mixture models (GMMs), deep unfolding (DU), lazy learning (LL), and so forth, have additionally been explored for FD SIC [119], [120], [122], [125], [126]. Integrating ML with FD communications has achieved considerable success in terms of performance and/or complexity when compared to the model-driven approaches. A comprehensive survey of such integration has thus far been lacking. Hence, this chapter addresses the knowledge gap in integrating data-driven ML approaches with FD communications, applying digital SIC.

The major contributions of this chapter are as follows:

- I firstly introduce a general and comprehensive system model to integrate ML with FD communications.
- I have briefly reviewed the traditional approaches for SIC in FD transceivers.
- I have surveyed the up-to-date contributions for applying data-driven ML approaches for SIC in FD systems, in which the SI can be directly learned from data rather than relying on traditional model-driven approaches.
- I have investigated the effect of using part of the output samples as features for training the SVR-based cancelers.
- I have provided a case study to assess the performance of the prominent ML approaches—in terms of SIC, power spectral density (PSD), training overhead, memory storage, and computational complexity—using two different test setups, i.e., two training datasets, and using various dataset sizes.
- I have devised an efficiency measure to select a suitable ML approach for SIC in FD transceivers, depending on the system requirements.
- I have highlighted the main challenges and potential research directions for successful adoption of ML approaches for canceling the SI in FD transceivers.

The rest of this chapter is organized as follows. Section 5.3 introduces the ML-based FD system model. Section 5.4 summarizes the traditional approaches for SIC in FD transceivers. Section 5.5 reviews the up-to-date contributions that apply ML approaches for SIC. Simulation results are presented in Section 5.6, challenges and future directions are summarized in Section 5.7, and finally, concluding remarks are drawn in Section 5.8.

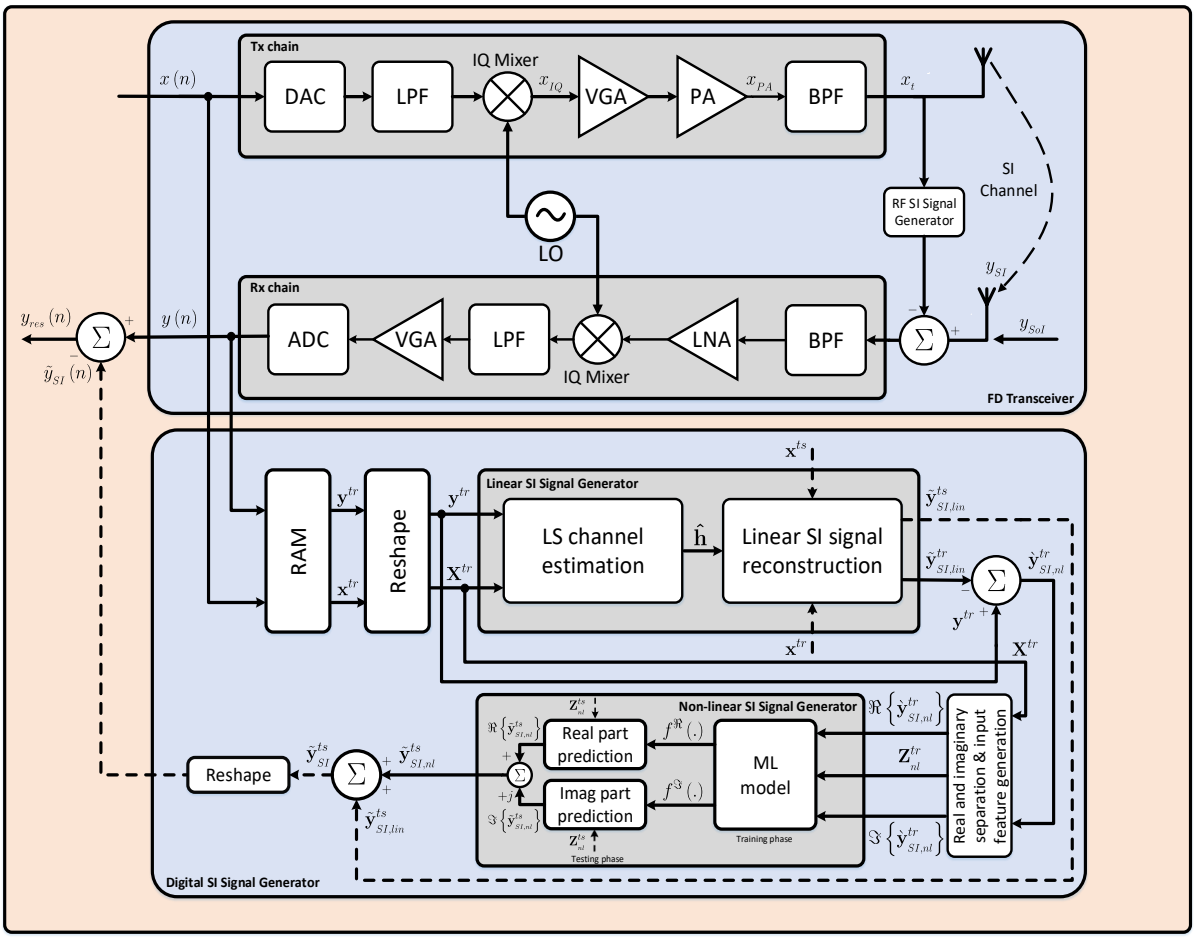


Fig. 5.1: ML-based FD system model with linear and non-linear digital cancellation stages.

5.3 ML-based FD System Model

The system model consisting of an FD transceiver with single transmit and single receive antennas, RF, and digital cancellation stages is illustrated in Fig. 5.1. At the Tx chain, the digital baseband samples, denoted by $x\,(n)$—with n as the sample index—are firstly distorted by the in-phase and quadrature-phase (IQ) imbalance of the mixer and then by the non-linearities of the PA. The digital equivalent of the baseband transmitted signal at the output of the Tx chain can be expressed as [99]-[101]

$$x_t(n) = \sum_{\substack{p=1,\\ p\,odd}}^{P} \sum_{m=0}^{M_{PA}} h_{m,p}\, x_{IQ}(n-m)^{\frac{p+1}{2}} x_{IQ}^{*}(n-m)^{\frac{p-1}{2}}, \tag{5.1}$$

with $x_{IQ}(n)$ as the IQ mixer's output signal and $(.)^*$ as the complex conjugate operator, whereas M_{PA}, $h_{m,p}$, and P are the memory depth, impulse response, and non-linearity order of the PA, respectively. In (5.1), p is an odd number, i.e., the odd-order non-linearities are only taken into account, e.g., $p \in \{3, 5..., 9\}$, as the even-order non-linearities are out-of-band and they are filtered by the Rx's analog and digital filters [100]. The transmitted signal x_t is propagated through an SI channel, forming an inevitable SI at the Rx chain. As a consequence, the received signal at the output of the Rx chain, i.e., at the output of the analog-to-digital converter (ADC), can be written as [127]

$$y\,(n) = y_{SI}\,(n) + y_{SoI}\,(n) + w\,(n)\,, \tag{5.2}$$

where $w\,(n) \sim \mathcal{CN}\,(0, \sigma^2)$ denotes the thermal noise, which is complex-valued Gaussian distributed with zero mean and variance σ^2, $y_{SoI}\,(n)$ indicates the received signal of interest (SoI), and $y_{SI}\,(n)$ represents the SI signal, which can be expressed as [99]-[101]

$$y_{SI}(n) = \sum_{\substack{p=1,\\ p\,odd}}^{P} \sum_{q=0}^{p} \sum_{m=0}^{M_i-1} h_{m,q,p}\, x\,(n-m)^q\, x^*\,(n-m)^{p-q}\,, \tag{5.3}$$

with $h_{m,q,p}$ as the impulse response of an overall channel containing the total effect of all transceiver impairments, e.g., PA non-linearities, IQ imbalance, and SI channel, and M_i as the memory effect introduced by the PA, SI channel delay spread at the Rx, etc.

To better evaluate the capabilities of the SI cancelers to suppress the SI signal

properly, I assume, for simplicity, that there is no SoI from any other FD transmit receive points (TRPs) and no mutual interference from any base station transmitting at the same frequency [90], [96], [97], [99]-[101]; hence, the received signal at the Rx chain's output will end up with the SI signal plus noise. The objective of the digital SI canceler is thus to suppress the SI to the Rx noise floor level. To that end, I firstly estimate the linear SI channel (i.e., causing the linear SI component) using the traditional LS channel estimation, which is performed for the case of single transmit and single receive antenna as follows [99]-[101]:

$$\hat{\mathbf{h}} = \left(\left(\mathbf{X}^{tr} \right)^{\mathrm{H}} \mathbf{X}^{tr} \right)^{-1} \left(\mathbf{X}^{tr} \right)^{\mathrm{H}} \mathbf{y}^{tr}, \tag{5.4}$$

with $(.)^{-1}$ and $(.)^{\mathrm{H}}$ as the inverse and conjugate transpose operators, respectively. The channel estimate $\hat{\mathbf{h}} \in \mathbb{C}^{M_i \times 1}$ while $\mathbf{X}^{tr} \in \mathbb{C}^{(N_{tr}-M_i)\times M_i}$, and $\mathbf{y}^{tr} \in \mathbb{C}^{(N_{tr}-M_i)\times 1}$ are respectively formed as

$$\mathbf{X}^{tr} = \begin{bmatrix} x(n) & x(n-1) & \cdots & x(n-M_i+1) \\ x(n+1) & x(n) & \cdots & x(n-M_i+2) \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ x(n+N_{tr}-M_i-1) & \cdots & \cdots & x(n+N_{tr}-2M_i) \end{bmatrix}. \tag{5.5}$$

and $\mathbf{y}^{tr} = [\ y(n) \quad y(n+1) \quad \cdots \quad y(n+N_{tr}-M_i-1)\ ]^{\mathrm{T}}$, with N_{tr} as the number of training samples and $(.)^{\mathrm{T}}$ as the transpose operator. Upon estimating the SI channel $\hat{\mathbf{h}}$, the linear SI component can be respectively reconstructed in the training and testing phases as follows:

$$\widetilde{\mathbf{y}}^{tr}_{SI,lin} = \hat{\mathbf{h}} \otimes \mathbf{x}^{tr}, \tag{5.6}$$

$$\widetilde{\mathbf{y}}_{SI,lin}^{ts} = \hat{\mathbf{h}} \otimes \mathbf{x}^{ts}, \tag{5.7}$$

where $\otimes$ indicates the convolution operator. $\mathbf{x}^{tr} \in \mathbb{C}^{(N_{tr}-M_i)\times 1}$ is formed from the training samples as $\mathbf{x}^{tr} = \begin{bmatrix} x(n) & x(n+1) & \cdots & x(n+N_{tr}-M_i-1) \end{bmatrix}^{\mathrm{T}}$, $\mathbf{x}^{ts} \in \mathbb{C}^{(N_{ts}-M_i)\times 1}$ is constructed similarly to $\mathbf{x}^{tr}$ from the testing samples (not from the training samples), and by replacing N_{tr} with N_{ts}, where N_{ts} represents the number of testing samples. Noting that, upon performing the convolution, the sequences $\widetilde{\mathbf{y}}_{\boldsymbol{SI,lin}}^{\boldsymbol{tr}}$ are resized to be aligned with the dimension of $\mathbf{y}^{tr}$.

The non-linear SI component, employed to train the ML approaches, e.g., NNs and SVRs, is obtained by subtracting the linear component from the original SI signal[2] as follows:

$$\grave{\mathbf{y}}_{SI,nl}^{tr} = \mathbf{y}^{tr} - \widetilde{\mathbf{y}}_{SI,lin}^{tr}. \tag{5.8}$$

Since the ML approaches are typically trained using real-valued inputs,[3] I separate the real and imaginary parts of $\mathbf{X}^{tr}$ and construct the input feature map, $\mathbf{Z}_{nl}^{tr} = \begin{bmatrix} \mathbf{z}(n) & \mathbf{z}(n+1) & \cdots & \mathbf{z}(n+N_{tr}-M_i-1) \end{bmatrix}^{\mathrm{T}}$, to train the non-linear canceler, with $\mathbf{z}(n) = \begin{bmatrix} \Re\{x(n)\} \ldots \Re\{x(n-M_i+1)\} & \Im\{x(n)\} \ \ldots \ \Im\{x(n-M_i+1)\} \end{bmatrix}$ for the case of the ML algorithms trained using the input samples only. However, for those trained with the input and output samples, $\mathbf{z}(n)$ will include a part of the output samples, as will be discussed later in Section 5.5. Upon constructing the input feature map, $\mathbf{Z}_{nl}^{tr}$, I separate the real and imaginary parts of $\grave{\mathbf{y}}_{SI,nl}^{tr}$ to serve as labels for training. Thus, during the training phase of the non-linear canceler, the input feature map $\mathbf{Z}_{nl}^{tr}$ is

[2]It is noted that training the ML approaches using the residual SI after linear cancellation, i.e., non-linear component, can enhance the SIC compared to the case when they are trained using both linear and non-linear components [97].

[3]Without loss of generality, it is assumed that training the ML approaches in the system model of Fig. 5.1 is done using real-valued inputs; however, complex-valued inputs can also be employed, as will be discussed in the following sections.

utilized with $\Re\left\{\grave{\mathbf{y}}_{SI,nl}^{tr}\right\}$ and $\Im\left\{\grave{\mathbf{y}}_{SI,nl}^{tr}\right\}$ to generate the modeling functions, $f^{\Re}(.)$ and $f^{\Im}(.)$, associated with approximating the real and imaginary parts of the non-linear SI signal, respectively. The real and imaginary parts can then be respectively predicted in the testing phase as

$$\Re\left\{\widetilde{\mathbf{y}}_{SI,nl}^{ts}\right\} = f^{\Re}\left(\mathbf{Z}_{nl}^{ts}\right), \tag{5.9}$$

$$\Im\left\{\widetilde{\mathbf{y}}_{SI,nl}^{ts}\right\} = f^{\Im}\left(\mathbf{Z}_{nl}^{ts}\right), \tag{5.10}$$

where $\mathbf{Z}_{nl}^{ts}$ is the non-linear canceler's testing matrix, which is formed similarly to $\mathbf{Z}_{nl}^{tr}$, but with replacing N_{tr} by N_{ts}. Based on the aforementioned, the non-linear SI signal is obtained by summing the real and imaginary parts as

$$\widetilde{\mathbf{y}}_{SI,nl}^{ts} = \Re\left\{\widetilde{\mathbf{y}}_{SI,nl}^{ts}\right\} + j\Im\left\{\widetilde{\mathbf{y}}_{SI,nl}^{ts}\right\}. \tag{5.11}$$

The estimated SI signal, i.e., after applying the linear and non-linear cancellations, can be expressed as

$$\widetilde{\mathbf{y}}_{SI}^{ts} = \widetilde{\mathbf{y}}_{SI,lin}^{ts} + \widetilde{\mathbf{y}}_{SI,nl}^{ts}, \tag{5.12}$$

and the residual SI signal can be written as

$$\mathbf{y}_{res}^{ts} = \mathbf{y}^{ts} - \widetilde{\mathbf{y}}_{SI}^{ts}. \tag{5.13}$$

The total SIC achieved upon applying the linear and non-linear cancellations can be quantified in dB as

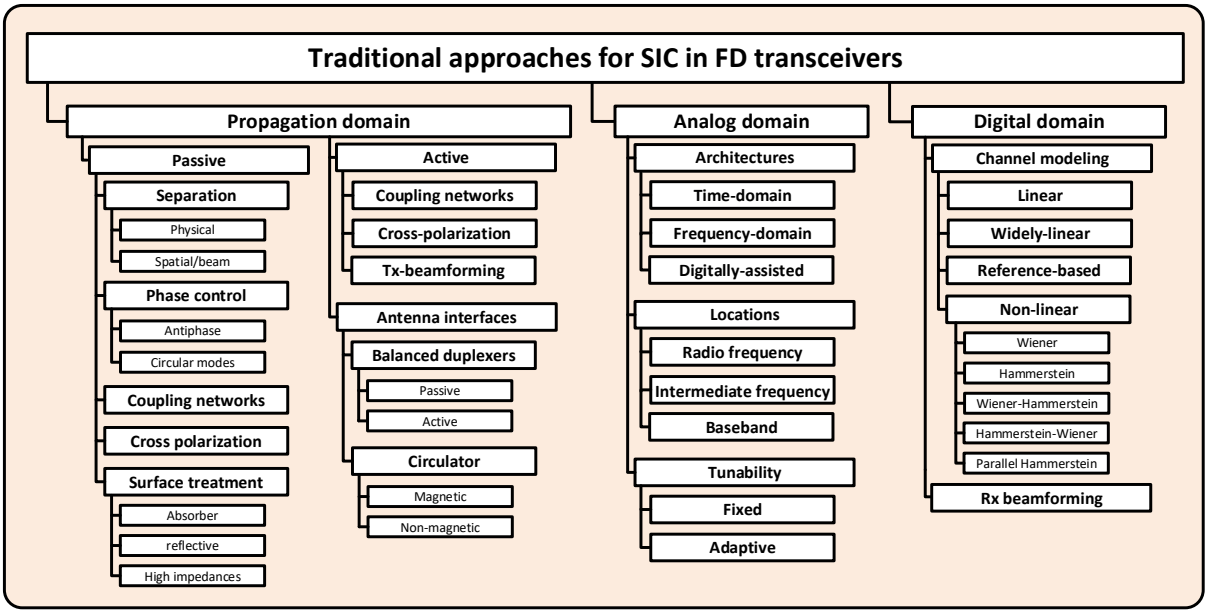


Fig. 5.2: Traditional approaches for SIC in FD transceivers [43].

$$\mathcal{C}_{dB} = 10\log_{10}\left(\frac{\sum_{n=1}^{N_{ts}}|y(n)|^2}{\sum_{n=1}^{N_{ts}}|y_{res}(n)|^2}\right), \tag{5.14}$$

with $y(n)$ and $y_{res}(n)$ as the n^{th} samples of $\mathbf{y}^{ts}$ and $\mathbf{y}^{ts}_{res}$, respectively.

5.4 Traditional Approaches for SIC in FD Transceivers

Canceling the SI in FD transceivers can be performed using various techniques that span the propagation, analog, and/or digital domains [28], [43], as summarized in Fig. 5.2. The following subsections briefly review such SIC approaches, discussing their advantages, disadvantages, and/or challenges.

5.4.1 Propagation Domain Self-Interference Cancellation

Canceling the SI within the propagation domain is typically performed at the early stage of the FD transceiver, i.e., it revolves around the Tx and Rx antennas. Propagation domain cancellation can be accomplished passively using techniques such as

antenna separation, coupling networks, phase control, cross-polarization, and/or surface treatments [28], [43], as shown in Fig. 5.2. Alternatively, it can be done actively using techniques such as active coupling networks, active cross-polarization, and/or Tx beamforming [43]. Additionally, antenna interfaces, such as balanced duplexers and circulators, can also be employed. Applying the SIC within the propagation domain has the advantage of refraining the SI signal from saturating the front end of the FD Rx; however, in some cases, it may lead to the suppression of the desired signal (i.e., SoI) [28]. Also, it can come at the cost of adding a hardware circuity to the FD transceiver. Hence, the focus is directed to additionally canceling the SI in other signal domains, e.g., analog and digital domains.

5.4.2 Analog Domain Self-Interference Cancellation

Canceling the SI within the analog domain is performed in the analog circuits between the antennas and digital conversion stages [28], [43]. Analog domain cancellation approaches have been classified based on their architecture, location, and tunability, as depicted in Fig. 5.2 [43]. One of the common architectures for analog domain cancellation is to use digitally-assisted techniques based on auxiliary transmit chains [43]. Digitally assisted analog domain cancellation has the advantage of preventing the SI signal from saturating the ADC, especially in mobility channel environments. However, it can result in an auxiliary transmit noise floor desensitization problem at the Rx. In addition to the Rx desensitization, the processing in the analog domain can be very costly and challenging to scale up into a higher number of antennas (i.e., multiple-input multiple-output (MIMO) scenario) [28]. The focus is thus directed to additionally canceling the SI in the digital domain, considering that the propagation and analog domain SIC have sufficient performance to provide the optimal dynamic range to the Rx's ADC.

5.4.3 Digital Domain Self-Interference Cancellation

Canceling the SI in the digital domain is performed after the ADC using techniques such as channel modeling and/or Rx beamforming, as shown in Fig. 5.2. Digital domain approaches, applying channel modeling techniques, use the fact that the Rx of any IBFD TRP has knowledge of its transmitted signal in order to model the transceiver's impairments. Specifically, in channel modeling-based SIC, linear, widely linear, and reference-based models are applied to approximate the SI channel effects. Additionally, non-linear models, such as Wiener, Hammerstein, Wiener-Hammerstein, and parallel Hammerstein, are employed to model the transceiver's non-linearities, as shown in Fig. 5.2. Digital domain cancellation has the advantage that the processing becomes relatively easy to perform and less hardware-expensive compared to the analog domain cancellation [28]; however, it can come at the cost of increasing the computational complexity of the FD transceiver [57].[4]

From the previous discussion, applying the traditional approaches for SIC in FD transceivers can come with challenges, such as imposing extra hardware, higher cost, and/or additional computational complexity. In contrast, applying the ML approaches for SIC in FD communications can relax such requirements, as reported in [90], [95]-[97], [99]-[101]. Given these potentials, more research efforts have recently been spurred to cancel the SI in FD transceivers using ML approaches. This chapter provides an in-depth survey of using the digital domain SIC based on ML non-linearity modeling techniques to tackle the SIC problem in FD transceivers.

[4] A detailed description of the traditional SIC approaches can be found in the survey papers [28] and [43].

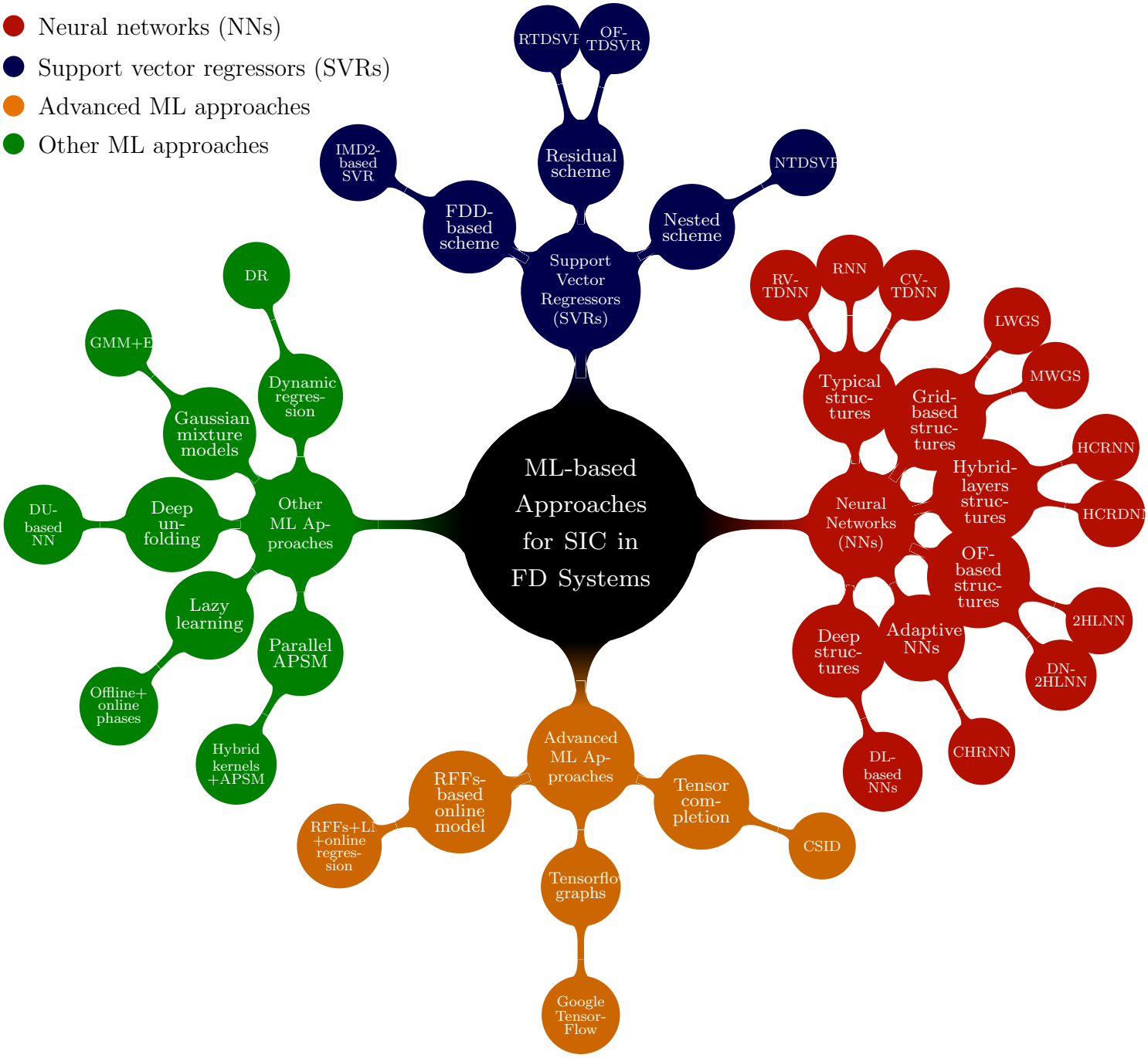


Fig. 5.3: ML-based approaches for SIC in FD transceivers.

5.5 ML-based Approaches for SIC in FD Transceivers

Fig. 5.3 summarizes the up-to-date contributions for applying ML-based approaches for SIC in FD transceivers. As can be seen from the figure, the SIC in FD systems can be performed using traditional ML approaches, such as NNs and SVRs. Also, advanced ML techniques, such as TC, TensorFlow graphs, and random Fourier features (RFFs), integrated with online learning, have been investigated for modeling the SI in FD transceivers. Other ML approaches, such as dynamic regression (DR), GMMs, DU, LL, and adaptive projected subgradient method (APSM), have also been studied for SIC. Among the different ML approaches applied for SIC, one can notice that NNs are the most popular due to their proven capabilities in modeling non-linearities with reduced complexity compared to other ML techniques. In this section, I aim to review and summarize the up-to-date research progress in applying ML-based approaches for SIC in FD transceivers.[5]

5.5.1 Neural Network (NN)-based SIC Approaches

Broadly speaking, canceling the SI in FD transceivers using ML mostly relies on NNs to make use of their potential compared to other ML approaches. As can be seen from the right-hand side of Fig. 5.3, a broad range of NN architectures, starting from typical NNs reaching to customized NN architectures, such as grid-based NNs, hybrid-layers NNs, and adaptive NNs, have been introduced for SIC in FD transceivers. The following subsections review and summarize the recent advances in applying NNs to SIC in FD transceivers.

[5]I use the term "advanced" to describe the recent—and not commonly utilized in other disciplines—ML approaches that are applied for SIC in FD transceivers. On the other hand, I employ the term "other" to describe the ML approaches—rather than the NNs and SVRs—that are frequently applied in other disciplines and subsequently introduced for SIC in FD transceivers.

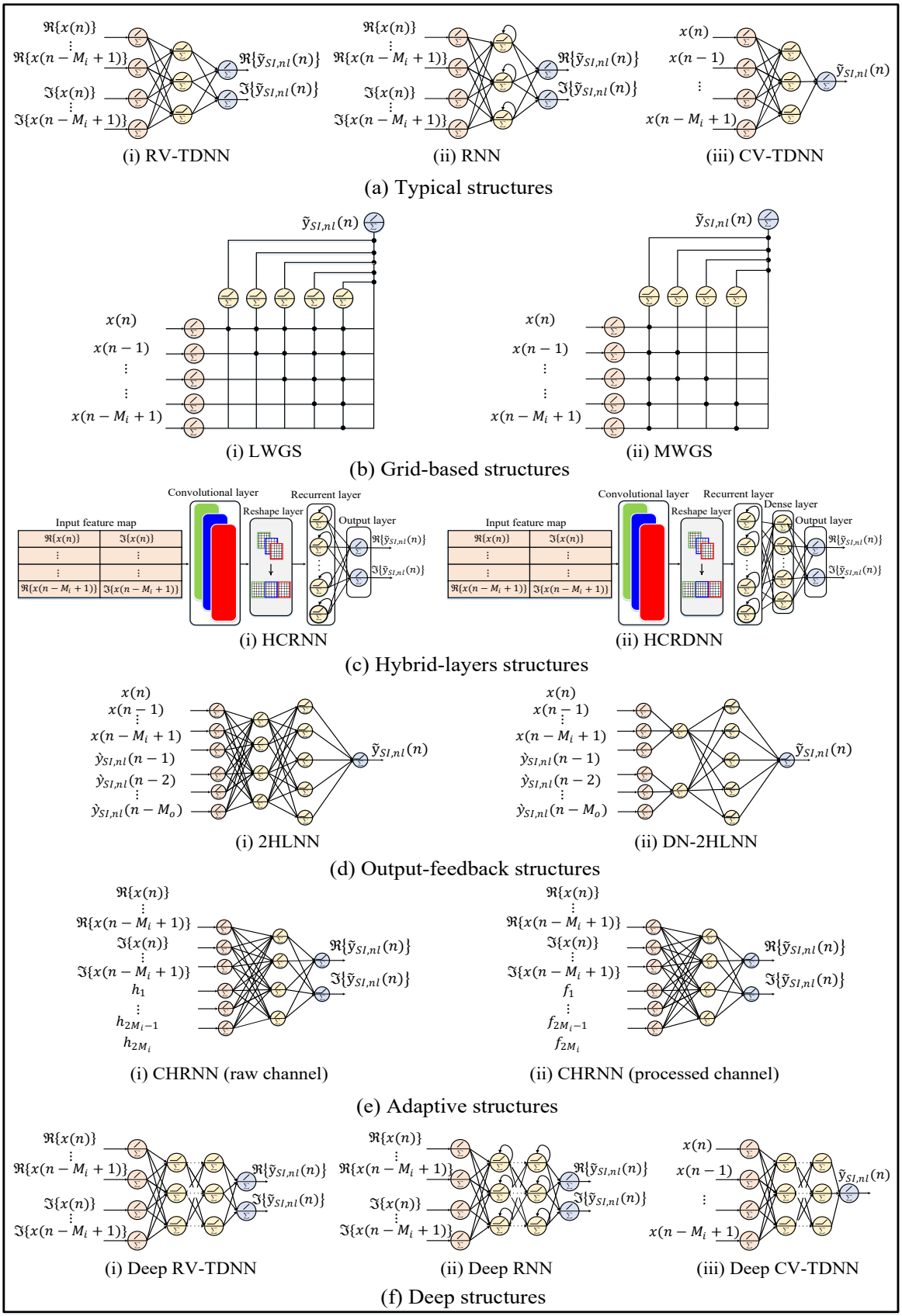


Fig. 5.4: NN-based approaches for SIC in FD transceivers.

5.5.1.1 Typical Structures

The first attempt to apply NNs for SIC in FD transceivers is done in [90], where a shallow feed-forward NN (FFNN) is utilized to approximate the non-linear SI signal. The FFNN in [90] is constructed—similarly to the real-valued time delay NN (RV-TDNN) in [128]—from an input layer fed by real-valued inputs consisting of current and past samples of the input signal to consider the FD system's memory effect.[6] The current and past samples are then transferred to a hidden layer to detect the FD system's non-linearities and finally to an output layer to estimate the target non-linear SI signal, as can be observed from Fig. 5.4(a-i). Simulation results show that the RV-TDNN could be beneficial from memory storage and computational complexity perspectives when compared to the polynomial model—a general form of the widely utilized parallel Hammerstein model [122]—at a similar SIC performance [90].[7] The hardware implementation of the NN-based cancelers is investigated in [95]-[96], where the RV-TDNN-based canceler is proved to be efficient in terms of area and energy consumption when compared to the polynomial-based canceler at a similar performance.

In [97], a typical recurrent NN (RNN) is introduced for canceling the interference in FD transceivers. The RNN [97] is trained similarly to the RV-TDNN using real-valued inputs consisting of current and past samples with memory. Contrary to the RV-TDNN [90], the RNN employs both forward and recurrent connections to enhance the learning capabilities [97], as can be seen from Fig. 5.4(a-ii). Applying a shallow RNN—with a single-hidden layer—for canceling the SI in FD transceivers can be beneficial from memory and computational complexity perspectives when compared to the typical RV-TDNN at a similar cancellation performance [97].

[6]Throughout this chapter, I will use the term RV-TDNN instead of FFNN.
[7]The RV-TDNN is also investigated for SIC in FD systems in [91]-[94].

In [97], [98], a complex-valued time delay NN (CV-TDNN) is investigated for canceling the FD system's SI. As can be observed from Fig. 5.4(a-iii), the CV-TDNN has a similar network architecture to that of RV-TDNN [90], while employing only one neuron instead of two neurons at the output layer. As its name implies, the CV-TDNN is trained using CV inputs and labels instead of the real-valued ones utilized in the case of RV-TDNN and RNN. Simulation results show that a shallow CV-TDNN-based canceler could be beneficial in terms of computational complexity when compared to its RV-TDNN and RNN counterparts at a similar SIC performance [90], [97].

5.5.1.2 Grid-based Structures

In [99], two grid-based NN structures, termed as ladder-wise grid structure (LWGS) and moving-window grid structure (MWGS), are introduced for modeling the interference in FD transceivers. The LWGS and MWGS are trained using CV data and built by a grid of connections—analog to nodes in the fully-connected NNs—between the input, hidden, and output layers' neurons, as shown in Fig. 5.4(b). As their names imply, the LWGS constructs the connections between the layers' neurons based on a ladder-wise topology, while the MWGS employs a moving window technique to arrange the connections, as can be seen from Figs. 5.4(b-i) and (b-ii), respectively. Using such a grid topology, the LWGS and MWGS exploit a fewer number of connections between the input and hidden layers' neurons to reduce the number of required parameters and, as a consequence, relax the computational complexity compared to the fully-connected NN counterparts. Simulation results indicate that the LWGS and MWGS [99] could achieve a comparable performance to that of CV-TDNN [97] while being beneficial in terms of memory storage and computational complexity.

5.5.1.3 Hybrid-Layers Structures

In [100], two hybrid-layers NN architectures—referred to as hybrid convolutional recurrent NN (HCRNN) and hybrid convolutional recurrent dense NN (HCRDNN)—have been introduced for learning the FD system's SI. The HCRNN and HCRDNN are trained using RV inputs and built using a combination of different NN layers, such as convolutional, recurrent, and dense layers, as shown in Fig. 5.4(c). The HCRNN and HCRDNN exploit the advantages of each layer in their network design to make use of their combined characteristics to improve the learning capabilities compared to the typical and grid-based NN architectures [90], [97], [99]. In particular, the HCRNN relies on a convolutional layer to use the weight-sharing property to reduce the number of required parameters and, consequently, relax the computational complexity. Further, it depends on a recurrent hidden layer to use its ability to learn the temporal behavior. On the other hand, the HCRDNN relies on an additional dense layer—added after the convolutional and recurrent layers—to build a highly predictive NN model with low computational complexity requirements. The HCRNN and HCRDNN [100] are shown to be beneficial from the computational complexity perspective while achieving a similar SIC performance compared to the typical and grid-based structures, albeit at the cost of increased memory requirements [90], [97], [99].

5.5.1.4 Output Feedback Structures

In [101], two output-feedback (OF)-based NN structures, namely two-hidden layers NN (2HLNN) and dual-neurons two-hidden layers NN (DN-2HLNN), have been introduced for canceling the SI in FD transceivers. As their names imply, the OF-based NN structures exploit a part of the output samples—fed back through a buffer to the input layer—to be utilized as features for training. In other words, the OF-based NN

structures are trained using an input feature map that considers not only the input samples as features for training but also the output samples, as shown in Fig. 5.4(d). Feeding part of the output samples for training helps to consider the effect of over-the-air SI propagation delay spread, which in turn enhances the learning capabilities, and as a consequence, improves the SIC performance compared to the NN structures only trained by the input samples. In the 2HLNN, a full connection is established between the input features—including both input and output samples—and the first hidden layer's neurons, as shown in Fig. 5.4(d-i). However, in the DN-2HLNN, the input features are not fully connected traditionally to the first hidden layer's neurons. The features related to the input samples are connected to one neuron to recognize the input signal's memory effect, while those related to the output samples are connected to another neuron to recognize the output signal's memory effect, as shown in Fig. 5.4(d-ii). Simulation results [101] reveal that the DN-2HLNN could be beneficial from memory storage and computational complexity perspectives while achieving a similar SIC performance to that of the LWGS, MWGS, HCRNN, HCRDNN, and 2HLNN [99], [100], [101].

5.5.1.5 Adaptive Structures

In [102], a channel adaptive NN structure, referred to as channel-robust NN (CHRNN), has been integrated with an LS-based linear canceler to model the SIC in FD transceivers over time-varying SI channels. In more detail, in [102], a linear canceler is trained continuously in each frame to estimate the channel coefficients, and a pre-trained NN is then utilized to construct the non-linear SI signal based on either raw or processed channel coefficients, as shown in Figs. 5.4(e-i) and (e-ii), respectively. For the former, the pre-trained NN is fed directly with the estimated channel coefficients obtained by the linear canceler, whereas for the latter, the pre-trained NN is fed

by a processed version of the estimated channel coefficients [102, Eq. (7)]. Simulation results indicate that CHRNN learns well when it is fed by processed channel coefficients rather than the raw ones. Further, the results reveal that the CHRNN-based canceler could lead to time reductions in computational complexity while attaining a similar performance to that of the polynomial-based canceler, adapted to handle time-varying SI channels [102].

5.5.1.6 Deep Structures

The concept of DL has also been introduced for modeling the interference in FD transceivers. In [97], deep versions of the typical RV-TDNN, RNN, and CV-TDNN, as shown in Fig. 5.4(f), have been introduced to model the SIC with lower memory and complexity. Using deep rather than shallow NNs is motivated by the fact that a deep NN with a small number of neurons in each layer, i.e., lower memory storage and computational complexity, can typically generalize better than a shallow NN with a large number of neurons in one layer [89]. Simulation results show that a deep CV-TDNN could be beneficial from memory storage and computational complexity perspectives while achieving a similar performance to that of a shallow CV-TDNN [97]. However, this is not applicable in all cases, as using a deep RNN increases the memory storage and computational complexity compared to the shallow RNN due to using many recurrent connections. Finally, adapting deep RV-TDNN for SIC results in decreasing the complexity while augmenting the memory storage compared to its shallow counterpart [97]. The concept of DL has also been studied for SIC in FD systems in other contexts, such as [103]-[105].

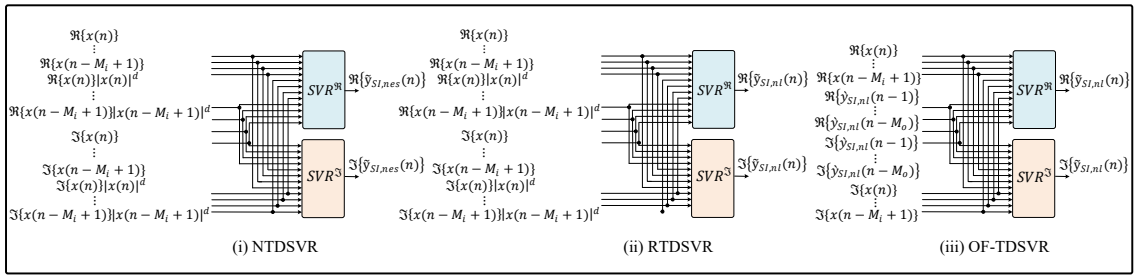


Fig. 5.5: SVR-based approaches for SIC in FD transceivers.

5.5.2 Support Vector Regressor (SVR)-based SIC Approaches

Despite being extensively used for SIC in FD transceivers, the NN-based cancelers are prone to some inherent characteristics of NN models, such as intolerable training complexity and less generalization when few examples are available for the training process. To overcome such bottlenecks, the SVRs, variants of support vector machines, have recently been introduced as alternatives to NNs for modeling the SI. The initial attempt of applying the SVRs for SIC is presented in [106] for frequency division duplex (FDD) transceivers—not for FD transceivers—where an SVR model is employed to generate a replica of the undesired transmit leakage-based second-order intermodulation distortion (IMD2) signal. Applying SVRs for SIC in FD systems came after in a few works in [107]-[108]. The following subsections review and summarize the few attempts to apply SVRs to cancel the SI in FD transceivers.

5.5.2.1 Nested-based Approaches

The first attempt to apply SVRs for SIC in FD transceivers is made in [107], where a non-linear time-delay SVR (TDSVR)-based canceler is integrated with a linear canceler—in a nested scenario—to suppress the SI signal down to the Rx noise floor level. The nested TDSVR (NTDSVR), shown in Fig. 5.5(i), is trained using an input feature map that considers the real and imaginary parts of the current and past input samples. Besides, the odd higher-order terms of the input samples (with memory) are also considered for training. The output labels for training the NTDSVR are created

by first estimating the SI channel; thereafter, an inverse filtering is applied to remove the effect of the linear SI channel [107]. Upon eliminating the channel effect, the output samples, denoted by $\tilde{y}_{SI,nes}$ in Fig. 5.5(i), and including the impact of non-linearity only, are served as labels to train the NTDSVR. After the non-linear SI component is reconstructed, the linear channel is then applied for linear component reconstruction. The estimated SI signal, including the linear and non-linear components, is then subtracted from the original SI signal to perform the SIC. Simulation results show that the NTDSVR-based canceler is beneficial in terms of SIC performance enhancement compared to the conventional non-linear polynomial-based cancelers [107].

5.5.2.2 Residual-based Approaches

a) *RTDSVR*: The second attempt to apply SVRs for SIC in FD transceivers is investigated in [108], where a residual-based TDSVR (RTDSVR) is introduced. The input feature map to train the RTDSVR is constructed similarly to the nested approach [107]. However, the output labels are created differently based on the residual output signal after applying the linear SIC, as can be seen from Fig. 5.5(ii). Particularly, in the residual scheme, the linear SI channel is first estimated, and then the linear SI signal's component is fully reconstructed. The estimated linear SI signal is then subtracted from the original SI signal, and the residual SI signal, denoted by $\tilde{y}_{SI,nl}$, and involving the non-linear component only, is utilized for training the RTDSVR. Upon reconstructing the non-linear SI, it is combined with the linear one before being subtracted from the original SI to perform the SIC. Simulation results reveal a superiority of the RTDSVR to improve the SIC compared to the NTDSVR, especially for low or moderate transmit power levels [108].[8]

[8]It is noted that the residual scheme applied for SVRs in [108] follows a similar mechanism to that of NNs-based cancelers, where the residual output signal's samples, after applying the linear SIC, are used as labels for training.

b) OF-TDSVR: Investigating the effect of feeding back part of the output samples to be exploited as features for training the SVR-based cancelers *has not previously considered in the literature and is examined for the first time in this chapter*, in which an SVR model, referred to as output-feedback time-delay SVR (OF-TDSVR), is integrated with a linear canceler in a residual scheme to suppress the SI signal. Similar to the OF-based NN structures, the OF-TDSVR is trained using an input feature map that considers both input and output samples as features for training, as shown in Fig. 5.5(iii). As proved for NNs, feeding part of output samples for training helps to consider the effect of over-the-air SI propagation delay spread, which in turn can enhance the learning capabilities and, subsequently, improve the SIC performance compared to the existing SVR-based cancelers—trained only by the input samples. Also, it may be beneficial for reducing the training overhead compared to the existing SVR and NN literature benchmarks.

5.5.3 Advanced ML-based SIC Approaches

Advanced ML approaches, such as TC, TensorFlow graphs, and RFFs, integrated with online learning, have recently been introduced for SIC in FD transceivers. The details of such advances are provided in the following subsections.

5.5.3.1 Tensor Completion (TC)

In [109], a canonical system identification (CSID) approach, based on a low-rank tensor constraint optimization problem, is utilized to approximate the non-linear SI signal as in the case of NNs and residual-based SVRs. In more detail, the CSID approach formulates the SIC problem as a low-rank tensor decomposition problem to be solved using an alternating least squares optimization algorithm. Simulation results [109] indicate that the CSID-based cancelers could achieve similar performance

to that of the polynomial and NN-based cancelers [90], [96]. Meanwhile, they can be beneficial from the computational complexity perspective at the cost of higher memory storage requirements.

5.5.3.2 TensorFlow Graphs

In [110], TensorFlow graphs, recent advances in ML, are introduced to cancel the SI in a real-time software-defined radio (SDR). Generally, graphs are exploited in ML to enable ML researchers/developers to write an abstracted version of their ML techniques in the form of data-flow graphs, which can then be utilized and applied to any of the ML algorithms [111]. Based on such graphs, in [110], the SIC is performed in real-time SDR based on an NN that employs a Google TensorFlow graph. Simulation results reveal that the TensorFlow graph-based approach could achieve a SIC that can reach the hardware limit and surpass existing digital non-ML-based SIC approaches in the literature [110].

5.5.3.3 Random Fourier Features (RFFs)

In [112], the RFFs and the least mean-squares (LMS) algorithm are integrated with online linear regression to perform the SIC in FD transceivers. Principally, RFFs are utilized to scale up kernel-based ML techniques by providing a non-linear transformation of input data to a higher dimensional feature space. So, non-linearities can be efficiently modeled using linear-based techniques in the original space, resulting in scalable, fastly-converged, and computationally efficient solutions [113], [114]. Based on this, in [112], the input samples are first transformed using RFFs, then the residual SI signal, after applying the linear SIC, is used with the transformed input to approximate the non-linear SI signal using an LMS-based canceler. The estimated signal is then subtracted from the original SI to obtain the residual SI signal; thereafter, an

estimation vector is updated online based on that residual and using an RFFs-based observation matrix. Simulation results show that an online RFFs-LMS-based canceler could be beneficial from SIC and complexity perspectives compared to batch learning algorithms involving NTDSVRs [112].[9]

5.5.4 Other ML-based SIC Approaches

Seeking more advantages in other ML approaches investigated in other disciplines, the DR, GMMs, DU, LL, and APSM have been explored for SIC in FD transceivers. The details of such approaches are provided in the following subsections.

5.5.4.1 Dynamic Regression (DR)

In [119], a classical DR model is introduced for canceling the interference in FD transceivers. Generally, DR models are exploited in ML problems to identify how related a certain output is to an input and allow future output forecasting. Based on this, in [119], a classical DR model is utilized to represent the memory effect caused by the amplifiers in FD systems. Upon estimating the DR coefficients, the SI signal is jointly estimated in time and frequency domains and is subtracted from the original SI signal to perform the digital SIC. Simulation results reveal that the DR-based SIC approach could achieve a high digital SIC performance and effectively attenuate the SI signal close to the Rx noise floor level. Besides, the DR-based SIC approach is validated using a real-time SDR platform and is able to properly provide a demonstration via video streaming [119].

[9] Although the RFFs are integrated with online regression in [112], they are utilized with various ML algorithms in other disciplines, such as [115]-[118].

5.5.4.2 Gaussian Mixture Models (GMMs)

In [120], an ML approach based on GMMs clustering is introduced to design an FD transceiver, which can detect the desired signal (i.e., SoI) directly without using digital-domain cancellation or even channel estimation. As the name implies, GMMs clustering uses a mixture, i.e., a superposition, of Gaussian distributions to fit the training data and assign the data points to a certain cluster based on their conditional probabilities [121]. In more detail, in [120], the received signal is firstly clustered, and a one-to-one mapping of the symbols, based on a GMMs clustering and an expectation-maximization (EM) algorithm, is utilized to perform the signal detection in each cluster. Simulation results reveal that an FD transceiver, utilizing the GMMs clustering, could achieve a comparable bit error rate with that of FD transceivers employing maximum likelihood detectors when perfect channel knowledge is considered and a better one when the LS/LMS channel estimation is used [120]. However, this transceiver is limited to operating scenarios when low-order modulation techniques are employed.

5.5.4.3 Deep Unfolding (DU)

In [122], an ML approach based on DU is introduced for canceling the interference in FD transceivers. DU involves converting the model-based methods, requiring iterative optimization algorithms for solving, into layer-wise structures analog to that of NNs [123], [124]. This enables fusing the iterative optimization methods with NNs' libraries/tools to cover a wide range of tasks and applications. The concept of DU is applied for SIC in [122], where a cascade of non-linear blocks—involving the impact of PA and IQ mixer non-linearities—is exploited with the traditional backpropagation algorithm to approximate the SI signal. Simulation results corroborate that the DU-based SIC approach could be beneficial from memory storage and compu-

tational complexity perspectives when compared to the literature benchmarks, e.g., polynomial- and CV-TDNN-based cancelers, at a similar SIC performance [122].

5.5.4.4 Lazy Learning (LL)

In [125], an ML approach based on LL is introduced to perform the SIC in cellular wireless networks operating with FD transmission. As their names imply, the LL-based models postpone the generalization to the training data until a system query is performed. Based on this concept, in [125], offline and online stages are exploited to generate the interference database and transmit the data, respectively. In the offline phase, the FD system's output signal excluding the SoI, is recorded in a database. However, in the online phase—in which the system is fully operated with the SoI—a suitable SI value is looked up in the offline-generated database with the help of a learning approach to perform the digital SIC. Simulation results show that the LL-based SIC approach could be effectively utilized for canceling the interference and enabling the FD transmission in cellular wireless networks [125].

5.5.4.5 Adaptive Projected Subgradient Method (APSM)

In [126], an ML SIC approach based on parallel APSM is introduced for canceling the interference in FD transceivers. Specifically, in [126], a hybrid kernel is first constructed by combining linear and non-linear Gaussian kernels. This kernel is then adapted to a parallel APSM approach where a non-linear function—approximating the SIC problem—is extracted using projection. Simulation results show that the hybrid kernel-based APSM approach could properly model the SI compared to a SIC method employing the normalized LMS filtering [126]. Moreover, it can also be parallelized, i.e., it can perform parallel processing to reduce the system latency.

Thus so far, I have surveyed the up-to-date contributions that apply ML-based

TABLE 5.1: Summary of ML-based approaches applied for SIC in FD transceivers.

ML approach	Key feature(s)	Methodology	Proposed Technique(s)	Ref.
Neural Networks	Typical structures	RV-TDNN	A typical RV-TDNN is introduced to learn the SI in FD transceivers with lower memory and complexity than those of model-based approaches, e.g., polynomial-based cancelers.	[90]-[94]
			A hardware implementation for NN-based cancelers, employing RV-TDNNs, is introduced.	[95], [96]
		RNN	A typical RNN is introduced to learn the SI in FD transceivers with lower memory and complexity than the polynomial and RV-TDNN-based cancelers.	[97]
		CV-TDNN	A typical CV-TDNN is introduced to learn the SI in FD transceivers with lower memory and complexity than the polynomial-, RV-TDNN-, and RNN-based cancelers.	[97], [98]
	Grid-based structures	LWGS	An NN structure based on a ladder-wise grid topology is introduced to learn the SI in FD transceivers with lower memory and complexity than the typical NN-based cancelers.	[99]
		MWGS	An NN structure based on a moving-window grid topology is introduced to learn the SI in FD transceivers with lower memory and complexity than the typical NN-based cancelers.	[99]
	Hybrid-layers structures	HCRNN	An NN structure based on convolutional and recurrent layers is introduced to learn the SI in FD transceivers with lower complexity than the typical and grid-based cancelers.	[100]
		HCRDNN	An NN structure based on convolutional, recurrent, and dense layers is introduced to learn the SI in FD transceivers with lower complexity than the typical and grid-based cancelers.	[100]
	Output-feedback structures	2HLNN	An NN structure based on feedback samples from the output layer is introduced to learn the SI in FD transceivers with lower memory and complexity than the typical, grid, and hybrid-layers NN-based cancelers.	[101]
		DN-2HLNN	An NN structure based on two neurons in the first hidden layer and feedback samples from the output layer is introduced to learn the SI in FD transceivers with lower memory and complexity than the typical, grid, and hybrid-layers NN-based cancelers.	[101]
	Adaptive structures	CHRNN	A channel adaptive NN structure, based on row channel input or processed channel input, is integrated with a linear canceler to learn the SI in FD transceivers with lower memory and complexity than the adaptive polynomial-based cancelers.	[102]
	Deep structures	DL-NNs	DL-based NN structures, employing multiple hidden layers, are introduced to learn the SI in FD transceivers with lower memory and complexity than the model-based approaches.	[103]-[105]
Support Vector Regressors	FDD-based scheme	IMD2-based SVR	An SVR-based canceler is introduced for canceling the IMD2 leakage signal in FDD, not in FD transceivers.	[106]
	Nested scheme	NTDSVR	An SVR-based canceler, using nested generated training labels, is introduced to learn the SI in FD transceivers, with lower training overhead than the NN-based cancelers.	[107]
	Residual scheme	RTDSVR	An SVR-based canceler, using residual generated training labels, is introduced to learn the SI in FD transceivers, with lower training overhead than the NN-based cancelers.	[108]
		OF-TDSVR	An SVR-based canceler, using residual generated training labels and feedback output samples, is introduced to learn the SI in FD transceivers with lower training overhead than the NN-based cancelers.	[This work]
Advanced ML Approaches	Tensor completion	CSID	A CSID approach based on a low-rank tensor decomposition problem is introduced to learn the SI in FD transceivers with lower complexity than the polynomial and NN-based cancelers.	[109]
	TensorFlow graphs	Google TensorFlow	Google TensorFlow graphs are integrated with a real-time SDR to cancel the SI with higher SIC than that of existing digital non-ML-based SIC approaches in the literature.	[110]
	Random Fourier features	RFFs+LMS+ online regression	An ML approach based on RFFs, LMS, and online regression is introduced to learn the SI in FD transceivers with lower complexity than the batch learning algorithms involving NTDSVRs.	[112]
Other ML Approaches	Dynamic regression	DR	A classical DR model is integrated with a real-time SDR to model the memory effects caused by amplifiers in FD transceivers.	[119]
	Gaussian mixture models	GMM+EM	An ML approach based on GMMs clustering is introduced for designing an FD transceiver that can detect the desired signal directly without using digital-domain cancellation or even channel estimation.	[120]
	Deep unfolding	DU-based NN	An ML approach based on DU, integrating a cascade of non-linear blocks to mimic the impact of PA and IQ mixer non-linearities, is introduced to learn the SI in FD transceivers with lower memory and complexity than the polynomial and NN-based cancelers.	[122]
	Lazy learning	Offline+online phases	An ML approach based on LL and integrating offline and online phases for generating the SI database and transmitting the data, respectively, is introduced to perform the SIC in FD-operated cellular wireless networks.	[125]
	Parallel APSM	Hybrid kernels+APSM	An ML approach based on a hybrid kernel—involving linear and non-linear Gaussian kernels—and parallel APSM is introduced to learn the SI in FD transceivers.	[126]

approaches for SIC in FD transceivers, as summarized in Table 5.1. The adaption of a particular ML-based approach for SIC depends on the system demands, such as the achieved SIC, training overhead, memory storage, and computational complexity. The following section will help to select a suitable ML-based approach for SIC in FD systems.

5.6 Simulation Results and Discussions

In this section, I provide a case study to compare the performance of the prominent ML approaches, surveyed in Section 5.5, with that of the polynomial canceler for two test setups (i.e., two training datasets) and using various dataset sizes. Specifically, I evaluate the prominent ML approaches in terms of the achieved SIC, PSD performance, training overhead, memory storage, and computational complexity and compare them with those of the polynomial-based canceler.

5.6.1 Selected Approaches

First, from the NN-based approaches shown on the right-hand side of Fig. 5.3, I select the typical NN architectures, i.e., RV-TDNN, RNN, and CV-TDNN [90], [97]; being the first literature benchmarks to apply ML approaches for SIC in FD transceivers. Further, I select the OF-based NN architectures, i.e., 2HLNN and DN-2HLNN, as proved to be efficient in terms of memory storage and computational complexity when compared to the other NNs [101]. Second, from the SVR-based approaches shown on the upper hand-side of Fig. 5.3, I select the RTDSVR [108] as it is shown to outperform the NTDSVR [107], especially for low or moderate transmit power levels. Additionally, I consider the investigated OF-TDSVR to be compared in reference to the existing NN and SVR benchmarks. Third and last, from the advanced and other

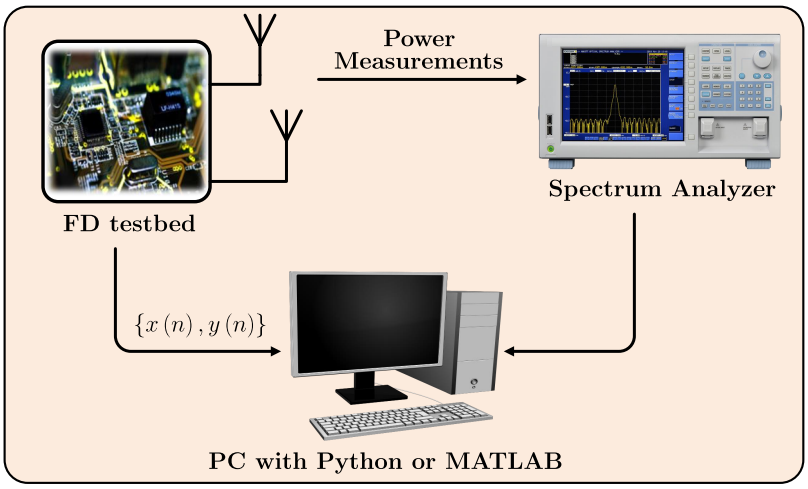


Fig. 5.6: Measurement setup.

ML approaches, shown on the lower- and left-hand sides of Fig. 5.3, I select the TC [109] and DU [122] approaches, as proven to be efficient in terms of memory storage and/or computational complexity when compared to the RV-TDNN and CV-TDNN, respectively. In the following subsections, I will evaluate and compare the previously selected approaches based on two test setups and using various performance metrics, such as the achieved SIC, PSD performance, training overhead, memory storage, and computational complexity.[10]

5.6.2 Measurement Setup

The measurement setup utilized to capture the datasets employed for training the prominent ML-based approaches selected in Section 5.6.1 is described in Fig. 5.6. Herein, an FD testbed, employing one transmit antenna and one receive antenna (1T1R), was set up in an indoor lab environment to generate two datasets [90], [96]. The first dataset [90] applies an orthogonal frequency division multiplexing (OFDM)

[10]Up to the author's knowledge, it is the first time in literature to compare the different ML-based SIC approaches based on two different test setups, i.e., two training datasets, and using various performance metrics, such as the SIC, PSD, training overhead, memory storage, and computational complexity.

TABLE 5.2: Measurement setup specifications.

Unit	Parameter	First dataset [90]	Second dataset [96]
FD testbed	Modulation	QPSK-modulated OFDM	QPSK-modulated OFDM
	FFT size	1024	2048
	Pass-band bandwidth	10 MHz	20 MHz
	Sampling frequency	20 MHz	80 MHz
	Average transmit power	10 dBm	32 dBm
	Passive analog suppression	53 dB	15 dB
	Active analog suppression	N/A	50 dB
	Total analog cancellation	53 dB	65 dB
	Transmit/receive antennas	1T1R	1T1R
	Dataset size	{2k, 3k, 4k, 5k}	{2k, 3k, 4k, 5k}
	Training/test splits	0.9/0.1	0.9/0.1
PC unit	Operating system	Windows 10	
	Processor	Intel(R) Xeon(R) W-2265, CPU @3.50GHz	
	# Cores	12	
	# Threads	24	
	RAM	128 GB	
	Python	3.7.5	
	Spyder	5.1.5	
	TensorFlow	2.0.0	
	Keras	2.3.1	
	NumPy	1.17.4	
	MATALB	R2020b	

signal with a quadrature phase-shift keying (QPSK) modulation and 10 MHz bandwidth, while the second [96] uses a QPSK-modulated OFDM signal with 20 MHz bandwidth. The average transmit power is set to 10 dBm and 32 dBm in the first and second datasets, respectively. The transmitted and received data are captured at 20 MHz and 80 MHz sampling rate for the first and second datasets, respectively. It is worth noting that using a higher sampling frequency enables the ML approaches to model the higher-order intermodulation distortion terms to efficiently suppress the SI, especially when high-transmit power levels are utilized.

At the Rx side of the FD testbed, total analog (i.e., passive and active) cancellations of 53 dB and 65 dB are applied in the first and second datasets, respectively, to

refrain the SI signal from saturating the FD-sensitive Rx chain. The digital received data after the ADC is then captured and retrieved to a personal computer (PC) for offline post-processing. In order to post-process the captured data at the PC, a 3.7.5 version of Python is installed in a Windows environment, using the 5.1.5 version of Spyder as the integrated environment for development, comparisons, and evaluation of different ML-based SIC approaches.[11] Finally, for analyzing the performance of various ML-based approaches at different dataset sizes, I have split each of the above-mentioned datasets into four separate datasets containing 2000, 3000, 4000, and 5000 samples, respectively. In all test cases, the first 90% of samples are used for training (and validation, if any), while the last 10% are reserved for testing. The specifications of the measurement setup employed in this work are detailed in Table 5.2.

5.6.3 Parameters Setting

The goal of this analysis is to find the peak performance of each SI canceler, e.g., polynomial, NN, SVR, TC, and DU. In other words, I aim to find the maximum SIC that each canceler can attain. Then, I compare the different cancelers in terms of the training overhead, memory storage, and computational complexity required to achieve their maximum SIC. To that extent: 1) for the polynomial canceler [90], I have optimized the non-linearity order P and memory length M_i; 2) for the NN-based cancelers, e.g., RV-TDNN, RNN, and CV-TDNN, etc. [90], [97], [101], I have optimized the memory length M_i along with the NN's hyperparameters, such as the number of hidden layers' neurons n_h, batch size (BS), learning rate (LR), activation function, and training optimizer; 3) for the SVR-based cancelers, i.e., RTDSVR and OF-TDSVR [108], I have obtained the optimum value for the memory length M_i, reg-

[11] It is noted that all ML-based SIC approaches selected for comparison in this chapter are implemented using Python integrated development environment, except for the TC, which is developed using the MATLAB independent development environment [109].

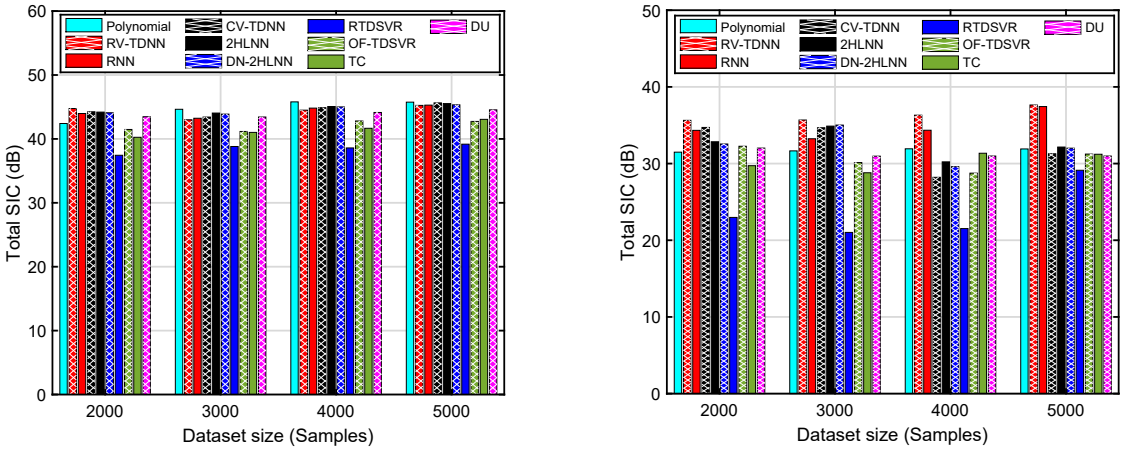
TABLE 5.3: Ranges for hyperparameters tuning for various SIC approaches.

ML Approach	Methodology	Hyperparameter	Hyperparameter tuning range
Model-based	Polynomial	Non-linearity order	$P \in \{3, 5..., 9\}$
		Memory length	$M_i \in \{2, 3..., 13\}$
NNs	Typical structures	Memory length	$M_i \in \{2, 3..., 13\}$
		Number of neurons	$n_{h1} \in \{2, 3..., 100\}$
		Learning rate	$LR \in \{0.5, 0.05, 0.005, 0.25, 0.025, 0.0025, 0.45, 0.045, 0.0045\}$
		Batch Size	$BS \in \{22, 62, 158, 256, 512, 1024\}$
		Activation function	$Act \in \{Relu, Sigmoid, tanh\}$
		Optimizer	$Opt \in \{Adam, RMSprop, SGD, Adadelta\}$
	OF structures	Memory length	$M_i \in \{2, 3..., 13\}$, $M_o = M_i - 1$
		Number of first layer neurons	$n_{h1} = 2$
		Number of second layer neurons	$n_{h2} \in \{2, 3..., 100\}$
		Learning rate	$LR \in \{0.5, 0.05, 0.005, 0.25, 0.025, 0.0025, 0.45, 0.045, 0.0045\}$
		Batch Size	$BS \in \{22, 62, 158, 256, 512, 1024\}$
		Activation function	$Act \in \{CRelu, AmpPhase, Cardioid, ModRelu\}$
		Optimizer	$Opt \in \{Adam, RMSprop, SGD, Adadelta\}$
SVRs	Residual schemes	Memory length	$M_i \in \{2, 3..., 13\}$, $M_o = M_i - 1$
		Regularization term	$C \in \{2^1, 2^2..., 2^7\}$
		Margin	$\epsilon \in \left\{\frac{1}{10^2}, ..., \frac{1}{10^3}, \frac{1}{4}, \frac{1}{4\times10}, ..., \frac{1}{4\times10^3}, \frac{1}{2}, \frac{1}{2\times10}, ..., \frac{1}{2\times10^3}\right\}$
		Gamma	$\gamma \in \left\{\frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1\right\}$
Advanced	Tensor completion	Memory length	$M_i \in \{2, 3..., 13\}$
		Tensor rank	$F \in \{1, 2..., 5\}$
		Quantization levels	$I \in \{4, 8..., 128\}$
		Regularization parameter	$\rho \in \{10^{-6}, 10^{-5}..., 10^{-3}\}$
		Smoothness factor	$\mu_n \in \{10^{-4}, 10^{-3}..., 10^{-1}\}$
Other	Deep unfolding	Memory length	$M_i \in \{2, 3..., 13\}$
		Learning rate	$LR \in \{0.5, 0.05, 0.005, 0.25, 0.025, 0.0025, 0.45, 0.045, 0.0045\}$
		Batch Size	$BS \in \{4, 6, ..., 10\}$

ularization term C, margin ϵ, along with the kernel hyperparameter, namely γ; 4) for the TC approach [109], I have tuned the memory length M_i, along with the optimization problem's hyperparameters, such as the tensor rank F, number of quantization levels I, regularization parameter ρ, and the smoothness factor μ_n; 5) for the DU approach, I have optimized the memory length M_i, and the LR and BS of the follow the regularized leader (FTRL) optimizer as in [122]. The ranges for hyperparameter tuning and the optimal values for hyperparameters over the first and second datasets are summarized in Tables 5.3 and 5.4, respectively.

TABLE 5.4: Optimal hyperparameters for different SIC approaches over the first and second datasets.

Data set	App.	Size	Method	Optimal Hyperparameters								Data set	App.	Size	Method	Optimal Hyperparameters							
				P	M_i	-	-	-	-	-	-					P	M_i	-	-	-	-	-	-
First dataset (implies lower transmit power, lower non-linear components, and narrower bandwidth)	Model-based	**2000**	Poly.	3	10	-	-	-	-	-	-	Second dataset (implies higher transmit power, higher non-linear components, and wider bandwidth)	Model-based	**2000**	Poly.	9	3	-	-	-	-	-	-
		3000		5	13	-	-	-	-	-	-			**3000**		9	4	-	-	-	-	-	-
		4000		7	13	-	-	-	-	-	-			**4000**		9	4	-	-	-	-	-	-
		5000		7	11	-	-	-	-	-	-			**5000**		9	6	-	-	-	-	-	-
	NNs	**Size**	**Method**	M_i	M_o	n_{h1}	n_{h2}	LR	BS	$Act.$	$Opt.$		NNs	**Size**	**Method**	M_i	M_o	n_{h1}	n_{h2}	LR	BS	$Act.$	$Opt.$
		2000	RV-TDNN	7	-	71	-	0.005	22	Relu	Adam			**2000**	RV-TDNN	3	-	96	-	0.005	22	Relu	Adam
			RNN	10	-	52	-	0.0045	256	Relu	Adam				RNN	5	-	100	-	0.005	158	Relu	Adam
			CV-TDNN	7	-	78	-	0.005	62	Cardioid	Adam				CV-TDNN	3	-	94	-	0.005	22	CRELU	Adam
			2HLNN	10	9	2	73	0.005	62	CRELU	Adam				2HLNN	7	6	2	100	0.025	158	CRELU	Adam
			DN-2HLNN	10	9	2	82	0.005	62	CRELU	Adam				DN-2HLNN	11	10	2	94	0.005	62	CRELU	Adam
		3000	RV-TDNN	12	-	48	-	0.005	22	Relu	Adam			**3000**	RV-TDNN	3	-	93	-	0.005	22	Relu	Adam
			RNN	9	-	60	-	0.0025	22	Sigmoid	Adam				RNN	5	-	94	-	0.0025	62	Relu	Adam
			CV-TDNN	8	-	53	-	0.025	158	Cardioid	Adam				CV-TDNN	13	-	69	-	0.0045	62	Cardioid	Adam
			2HLNN	10	9	2	51	0.0045	62	CRELU	Adam				2HLNN	10	9	2	100	0.005	62	Cardioid	Adam
			DN-2HLNN	11	10	2	100	0.0045	62	CRELU	Adam				DN-2HLNN	11	10	2	92	0.005	62	Cardioid	Adam
		4000	RV-TDNN	9	-	76	-	0.005	22	Relu	Adam			**4000**	RV-TDNN	4	-	97	-	0.0025	22	Relu	Adam
			RNN	9	-	45	-	0.0025	22	Sigmoid	Adam				RNN	5	-	98	-	0.0025	62	Relu	Adam
			CV-TDNN	13	-	36	-	0.0045	22	Cardioid	Adam				CV-TDNN	8	-	38	-	0.0045	256	Cardioid	Adam
			2HLNN	11	10	2	66	0.0045	62	CRELU	Adam				2HLNN	9	8	2	11	0.0045	62	Cardioid	Adam
			DN-2HLNN	10	9	2	94	0.0045	62	Cardioid	Adam				DN-2HLNN	6	5	2	10	0.0045	62	CRELU	Adam
		5000	RV-TDNN	9	-	57	-	0.005	22	Relu	Adam			**5000**	RV-TDNN	6	-	100	-	0.0025	22	Relu	Adam
			RNN	10	-	34	-	0.0025	158	Relu	Adam				RNN	6	-	90	-	0.0025	62	Relu	Adam
			CV-TDNN	11	-	30	-	0.0045	22	Cardioid	Adam				CV-TDNN	12	-	89	-	0.0045	158	CRELU	Adam
			2HLNN	11	10	2	96	0.0045	62	CRELU	Adam				2HLNN	12	11	2	81	0.0045	62	CRELU	Adam
			DN-2HLNN	13	12	2	65	0.0045	62	CRELU	Adam				DN-2HLNN	11	10	2	76	0.0045	62	CRELU	Adam
	SVRs	**Size**	**Method**	M_i	M_o	$C^{\Re}$	$\epsilon^{\Re}$	$\gamma^{\Re}$	$C^{\Re}$	$\epsilon^{\Re}$	$\gamma^{\Re}$		SVRs	**Size**	**Method**	M_i	M_o	$C^{\Re}$	$\epsilon^{\Re}$	$\gamma^{\Re}$	$C^{\Re}$	$\epsilon^{\Re}$	$\gamma^{\Re}$
		2000	RTDSVR	7	-	2	0.1	0.0625	2	0.1	0.0625			**2000**	RTDSVR	13	-	4	0.01	0.0625	4	0.0001	0.0625
			OF-TDSVR	9	8	8	0.01	0.0625	16	0.0001	0.0625				OF-TDSVR	3	2	64	0.025	0.0625	128	0.05	0.0625
		3000	RTDSVR	7	-	2	0.1	0.0625	2	0.1	0.0625			**3000**	RTDSVR	12	-	4	0.01	0.0625	4	0.01	0.0625
			OF-TDSVR	7	6	16	0.1	0.0625	16	0.1	0.0625				OF-TDSVR	5	4	32	0.0025	0.0625	64	0.05	0.0625
		4000	RTDSVR	11	-	2	0.1	0.0625	2	0.0001	0.0625			**4000**	RTDSVR	3	-	4	0.005	0.0625	8	0.0001	0.0625
			OF-TDSVR	7	6	8	0.1	0.0625	8	0.1	0.0625				OF-TDSVR	5	4	32	0.01	0.0625	64	0.05	0.0625
		5000	RTDSVR	7	-	2	0.25	0.0625	1	0.025	0.0625			**5000**	RTDSVR	12	-	4	0.005	0.0625	4	0.01	0.0625
			OF-TDSVR	9	8	32	0.1	0.0625	8	0.05	0.0625				OF-TDSVR	5	4	32	0.05	0.0625	64	0.05	0.0625
	Advanced	**Size**	**Method**	M_i	M_o	F	I	ρ	μ_n	-	-		Advanced	**Size**	**Method**	M_i	M_o	F	I	ρ	μ_n	-	-
		2000	TC	13	-	3	16	0.00001	0.1	-	-			**2000**	TC	2	-	4	32	0.0001	0.001	-	-
		3000		13	-	5	8	0.0001	0.01	-	-			**3000**		2	-	5	32	0.000001	0.001	-	-
		4000		13	-	5	16	0.0001	0.01	-	-			**4000**		2	-	5	32	0.00001	0.001	-	-
		5000		13	-	5	16	0.0001	0.01	-	-			**5000**		2	-	5	64	0.00001	0.001	-	-
	Other	**Size**	**Method**	P	M_i	M_o	LR	BS	$Opt.$	-	-		Other	**Size**	**Method**	P	M_i	M_o	LR	BS	$Opt.$	-	-
		2000	DU	5	9	-	0.25	4	FTRL	-	-			**2000**	DU	9	12	-	0.25	4	FTRL	-	-
		3000		5	9	-	0.05	6	FTRL	-	-			**3000**		7	12	-	0.25	6	FTRL	-	-
		4000		5	9	-	0.045	4	FTRL	-	-			**4000**		7	5	-	0.25	4	FTRL	-	-
		5000		5	13	-	0.25	6	FTRL	-	-			**5000**		9	5	-	0.25	6	FTRL	-	-



(a) First dataset.

(b) Second dataset.

Fig. 5.7: SIC of different ML-based SI cancelers compared to the polynomial canceler over the first and second datasets.

5.6.4 Performance Comparison

In this subsection, I assess the performance of the prominent ML-based SIC approaches in terms of their SIC, PSD, training time, memory storage, and computational complexity and compare them with those of the polynomial model. Afterward, I evaluate the efficiency of each canceler according to system demands. All the SIC approaches considered in this analysis are trained using the datasets described in Section 5.6.2, and with parameter settings optimized in Section 5.6.3.

5.6.4.1 SIC Performance

The total SIC achieved by different ML-based SIC approaches compared to the polynomial model upon tested using the first and second datasets, and with 2000, 3000, 4000, and 5000 samples is shown in Figs. 5.7a and 5.7b, respectively.[12] From the

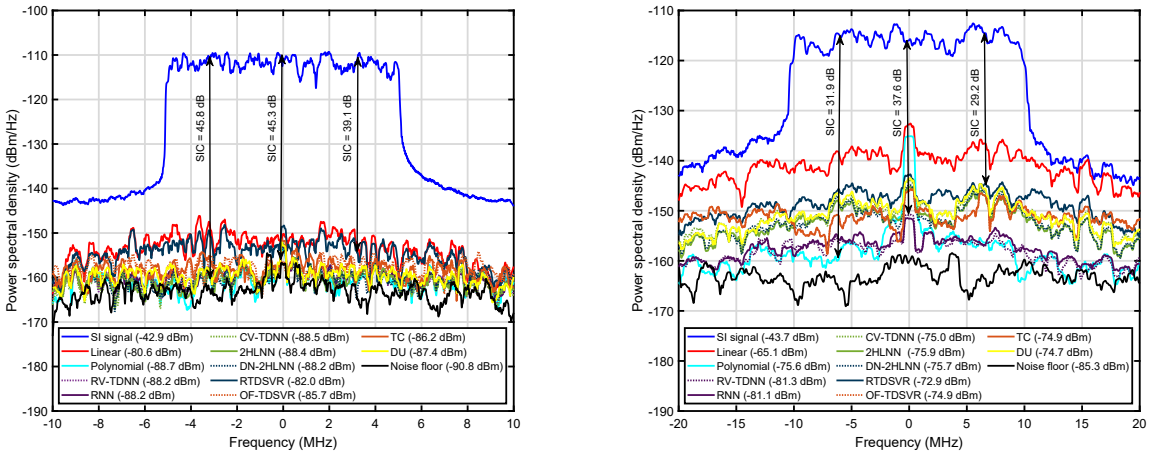
[12]In this chapter, I provide a case study to compare the performance of different ML approaches with the polynomial canceler when achieving the maximum SIC (i.e., peak-performance) at short dataset sizes, e.g., 2000, 3000, 4000, and 5000 samples. However, in my previous works in [99], [100], [101], I have compared the different ML approaches with the polynomial canceler when attaining a similar SIC (i.e., equi-performance) at a large dataset size, e.g., 20,000 samples. Accordingly, some of the results obtained in this chapter may differ from those reported in [99], [100], [101].

TABLE 5.5: SIC of different approaches when trained using 5000 samples of the first and second datasets.

Dataset size	Method	First dataset			Second dataset		
		Lin. Canc.	Non-lin. Canc.	Total Canc.	Lin. Canc.	Non-lin. Canc.	Total Canc.
5000 samples	Poly.	37.8	8.0	45.8	21.4	10.5	31.9
	RV-TDNN	37.8	7.5	45.3	21.4	16.3	37.6
	RNN	37.7	7.6	45.3	21.4	16.0	37.4
	CV-TDNN	37.7	7.9	45.6	21.4	9.90	31.3
	2HLNN	37.7	7.8	45.5	21.4	10.8	32.2
	DN-2HLNN	37.7	7.6	45.3	21.4	10.6	32.0
	RTDSVR	37.7	1.4	39.1	21.4	7.8	29.2
	OF-TDSVR	37.8	5.0	42.8	21.4	9.8	31.2
	TC	37.5	5.8	43.3	19.1	12.1	31.2
	DU	NA	NA	44.5	NA	NA	31.0

figures, one can observe that in the first dataset, where a low average transmit power is employed, the polynomial-based canceler achieves the highest cancellation performance compared to other cancelers for most of the dataset sizes. However, in the second dataset, where a high average transmit power is utilized, the RV-TDNN-based canceler provides the highest cancellation among the other cancelers for all dataset sizes. It can also be inferred from the figures that the RTDSVR achieves the lowest cancellation performance among the others, even if a low or high transmit power is utilized. Further, one can notice that employing a part of the output samples as features for training the SVR models can enhance the cancellation performance compared to the existing RTDSVR, i.e., the OF-TDSVR attains a significantly higher SIC than the RTDSVR benchmark. In sum, the polynomial canceler could be a good choice when a low transmit power is utilized, i.e., low transmit power generates less non-linearity SI signals. However, when a higher transmit power is employed, the RV-TDNN could be a better choice, i.e., high transmit power generates higher non-linearity SI signals.[13]

[13]Although all SI cancelers achieve a high non-linear cancellation in the second dataset compared to that attained in the first, as a result of having increased non-linearity, I interestingly note that the total SIC achieved in the former is lower than that in the latter, as can be seen from the sample results in Table 5.5. This is due to the degradation of the linear canceler's performance with increased non-linearity.



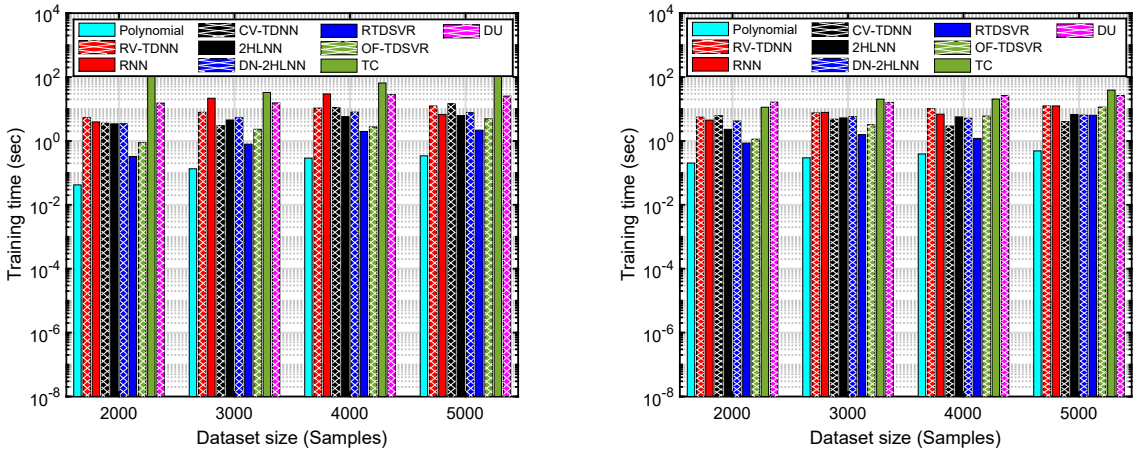
(a) First dataset.

(b) Second dataset.

Fig. 5.8: PSD of different ML-based SI cancelers compared to the polynomial canceler over the first and second datasets.

5.6.4.2 PSD Performance

The power spectra of the residual SI signal after applying the different ML-based SIC approaches compared to that of the polynomial-based canceler when tested using the first and second datasets and with 5000 samples, as an example, is shown in Figs. 5.8a and 5.8b, respectively. From Fig. 5.8a, one can observe that the polynomial-based canceler is able to suppress the SI signal with the lowest gap to Rx noise floor among the other cancelers in the first dataset; it can provide a gap to Rx noise floor value of $(90.8 - 88.7 = 2.1$ dB), bringing the SI signal very close to the Rx noise floor level. It can also be inferred from Fig. 5.8b that the RV-TDNN-based canceler provides the lowest gap to Rx noise floor compared to the others in the second dataset; it attains a gap to Rx noise floor value of $(85.3 - 81.3 = 4$ dB), bringing the SI signal close to the Rx noise floor level. The low gap to Rx noise floor achieved by the RV-TDNN compared to the polynomial canceler in the second dataset comes from the fact that it can reduce the leakage of the carrier around the DC tone, as shown in Fig. 5.8b



(a) First dataset. (b) Second dataset.

Fig. 5.9: Training time of different ML-based SI cancelers compared to the polynomial canceler over the first and second datasets.

[96]. Finally, one can observe from the figures that the SIC values achieved by the polynomial, RV-TDNN, and RTDSVR cancelers, as an example, match those reported in Table 5.5.

5.6.4.3 Training Overhead

In this subsection, I assess the training time, i.e., fitting time, required by each SI canceler to complete the training process. Specifically, for the polynomial-based canceler, I evaluate the training time needed to estimate the polynomial model's coefficients based on the LS algorithm. For the NN- and DU-based cancelers, I calculate the training time as the average training time required over different random seeds. For the SVR models, I approximate the training time as the maximum between the times needed to fit the $SVR^{\Re}$ and $SVR^{\Im}$, associated with estimating the real and imaginary parts of the non-linear SI signal, respectively, as shown in Fig. 5.5. Finally, for the TC-based canceler, I evaluate the training time required for fitting the low-rank tensor decomposition problem. Based on the aforementioned, in Figs. 5.9a and 5.9b,

I depict the training time of all the ML-based cancelers compared to the polynomial model upon tested using the first and second datasets, respectively. From the figures, it can be observed that the polynomial-based canceler requires the lowest training time among the others even if low or high average transmit power is employed, i.e., even if it is trained using the first or second dataset. Further, one can notice that the RTDSVR shows a good training time, i.e., it requires a lower training time than all other cancelers except the polynomial-based canceler. One can also observe that the SIC enhancement provided by the OF-TDSVR comes at the cost of increasing its training time compared to the RTDSVR benchmark. Additionally, it can be noticed that the TC- and DU-based cancelers require significantly higher training than the others, making them unfavorable choices for SIC, especially for operating scenarios where the training time is of interest. Finally, it can be observed from the figures that typically, as the dataset size increases, the training time of all SI cancelers increases as well.

5.6.4.4 Memory Storage

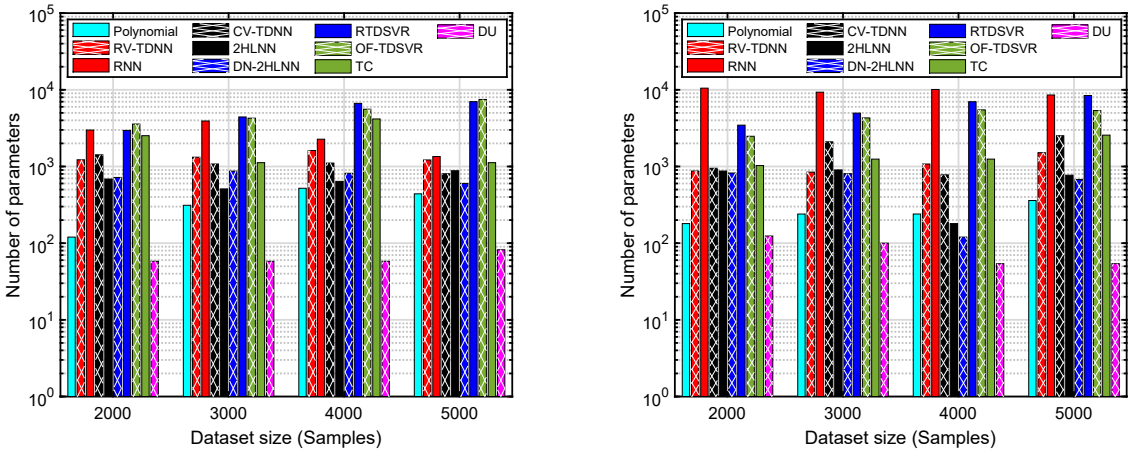
In this subsection, I assess the memory storage of different ML approaches in terms of the total number of parameters required in the inference stage and compare it with that of the polynomial model. Specifically, the number of parameters of the polynomial-based canceler is calculated as $2M_i + 2M_i \left\{ \left(\frac{P+1}{2} \right) \left(\frac{P+1}{2} + 1 \right) - 1 \right\}$ [90]. Further, the number of parameters of the typical RV-TDNN, RNN, and CV-TDNN is respectively evaluated as $2M_i \left(n_h + 1 \right) + 3n_h + 2$, $2M_i + n_h \left(n_h + 5 \right) + 2$, and $2M_i + 2 \left(M_i n_h + 2n_h + 1 \right)$, with n_h as the number of hidden neurons [90], [97]. The number of parameters of the OF-based NN structures, i.e., 2HLNN and DN-2HLNN, is respectively calculated as $2M_i + 2 \left\{ n_{h1} \left(M_i + M_o + n_{h2} + 1 \right) + 2n_{h2} + 1 \right\}$, and $2M_i + 2 \left(M_i + M_o + 4n_{h2} + 3 \right)$, with n_{h1} and n_{h2} as the number of neurons in the first and

TABLE 5.6: Memory storage and computational complexity of different SIC approaches.

Method	Number of parameters	Number of FLOPs
Polynomial [90]	$2M_i + 2M_i\left\{\left(\frac{P+1}{2}\right)\left(\frac{P+1}{2}+1\right)-1\right\}$	$10M_i + 10M_i\left\{\left(\frac{P+1}{2}\right)\left(\frac{P+1}{2}+1\right)-1\right\}-2$
RV-TDNN [90]	$2M_i\left(n_h+1\right)+3n_h+2$	$10M_i + n_h\left(4M_i+5\right)$
RNN [97]	$2M_i + n_h\left(n_h+5\right)+2$	$10M_i + 2n_h\left(n_h+\frac{9}{2}\right)$
CV-TDNN [97]	$2M_i + 2\left(M_i n_h + 2n_h + 1\right)$	$10\left\{M_i\left(n_h+1\right)+\frac{6}{5}n_h\right\}$
2HLNN [101]	$2M_i + 2\left\{n_{h1}\left(M_i+M_o+n_{h2}+1\right)+2n_{h2}+1\right\}$	$10M_i + 10\left\{n_{h1}\left(M_i+M_o\right)+n_{h1}n_{h2}+\frac{6}{5}n_{h2}\right\}$
DN-2HLNN [101]	$2M_i + 2\left(M_i+M_o+4n_{h2}+3\right)$	$10M_i + 10\left(M_i+M_o+\frac{16}{5}n_{h2}\right)$
RTDSVR [108]	$2M_i + N_{sv}^{\Re} + N_{sv}^{\Im} + 8$	$10M_i + 4dM_i\left(\frac{N_{sv}^{\Re}+N_{sv}^{\Im}}{2}\right)Q$
OF-TDSVR	$2M_i + N_{sv}^{\Re} + N_{sv}^{\Im} + 8$	$10M_i + 4d\left(M_i+M_o\right)\left(\frac{N_{sv}^{\Re}+N_{sv}^{\Im}}{2}\right)Q$
TC [109]	$2\left\{M_i\left(2FI+1\right)\right\}$	$8M_i\left(2F+1\right)-3F-7$
DU [122]	$2\left\{M_i\left(\frac{P+1}{2}\right)+2\right\}$	$10M_i\left(\frac{P+1}{2}\right)+P+18$

second hidden layers, respectively [101]. The number of parameters of the SVR models, i.e., RTDSVR and OF-TDSVR, employing a radial basis function (RBF) kernel, is evaluated as $2M_i + N_{sv}^{\Re} + N_{sv}^{\Im} + 8$, with $N_{sv}^{\Re}$ and $N_{sv}^{\Im}$ as the number of support vectors required to approximate the unknown functions of $SVR^{\Re}$ and $SVR^{\Im}$, respectively [108], [129]. Finally, the number of parameters for the TC- and DU-based cancelers is respectively given by $2\left\{M_i\left(2FI+1\right)\right\}$ and $2\left\{M_i\left(\frac{P+1}{2}\right)+2\right\}$, with F and I indicating the tensor rank and the number of quantization levels employed in the TC approach, respectively [109], [122]. A summary of the total parameters utilized to evaluate the memory storage of various SI cancelers is shown in Table 5.6.

Based on the aforementioned, I depict the number of parameters required by the various SI cancelers when tested by the first and second datasets in Figs. 5.10a and 5.10b, respectively. From the figures, one can observe that the DU-based canceler requires the lowest number of parameters compared to the others for both datasets and for all dataset sizes. The SVR-based cancelers, i.e., RTDSVR and OF-TDSVR, require the highest number of parameters among the others in the first dataset, as their parameters basically depend on the number of support vectors, i.e., $N_{sv}^{\Re}$ and $N_{sv}^{\Im}$, which in turn depend on the number of training data [129]. Thus, one can



(a) First dataset.

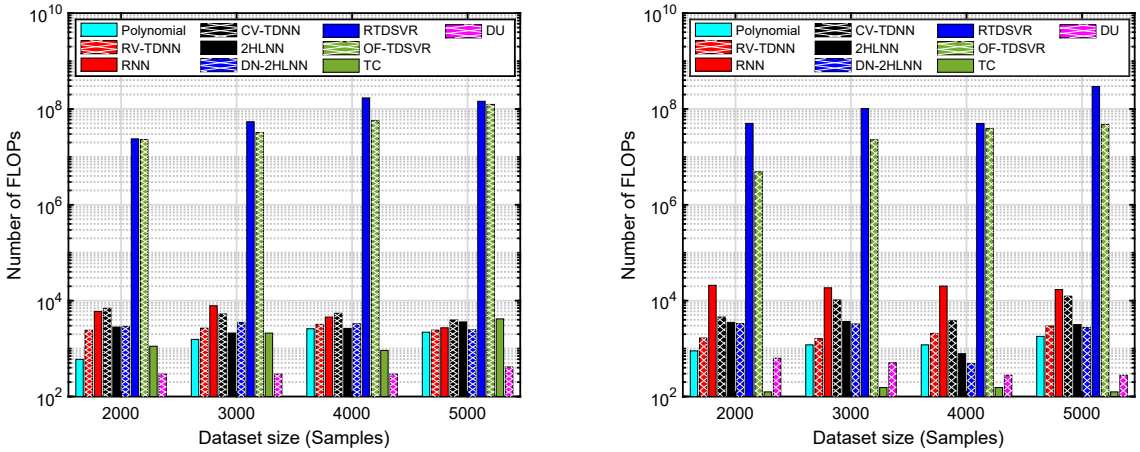
(b) Second dataset.

Fig. 5.10: Memory storage of different ML-based SI cancelers compared to the polynomial canceler over the first and second datasets.

notice from Figs. 5.10a and 5.10b that as the dataset size increases, the SVR models' parameters significantly increase as well. Finally, it can be inferred from the figures that the RNN-based canceler requires the highest number of parameters compared to the others in the second dataset as a result of using many recurrent connections.

5.6.4.5 Computational Complexity

In this subsection, I evaluate the computational complexity of various ML-based SIC approaches in terms of the total number of floating-point operations (FLOPs) required in the inference stage and compare it with that of the polynomial model. Particularly, the number of FLOPs of the polynomial-based canceler is calculated as $10M_i + 10M_i\left\{\left(\frac{P+1}{2}\right)\left(\frac{P+1}{2}+1\right)-1\right\}-2$ [90]. Besides, the number of FLOPs of the typical RV-TDNN, RNN, and CV-TDNN are respectively evaluated as $10M_i + n_h\left(4M_i+5\right)$, $10M_i + 2n_h\left(n_h+\frac{9}{2}\right)$, and $10\left\{M_i\left(n_h+1\right)+\frac{6}{5}n_h\right\}$ [90], [97]. Further, the number of FLOPs of the 2HLNN and DN-2HLNN are calculated as $10M_i + 10\left\{n_{h1}\left(M_i+M_o\right)+n_{h1}n_{h2}+\frac{6}{5}n_{h2}\right\}$ and $10M_i + 10\left(M_i+M_o+\frac{16}{5}n_{h2}\right)$, respectively



(a) First dataset. (b) Second dataset.

Fig. 5.11: FLOPs of different ML-based SI cancelers compared to the polynomial canceler over the first and second datasets.

[101]. On the other hand, the number of FLOPs of the SVR models, i.e., RTDSVR and OF-TDSVR, employing an RBF kernel, are respectively evaluated in the worst case as $10M_i+4dM_i\left(\frac{N_{sv}^{\Re}+N_{sv}^{\Im}}{2}\right)Q$ and $10M_i+4d\left(M_i+M_o\right)\left(\frac{N_{sv}^{\Re}+N_{sv}^{\Im}}{2}\right)Q$, with d and Q as the degree (e.g., $d=3$ for RTDSVR and $d=1$ for OF-TDSVR) and the number of testing samples, respectively [108]. Finally, the number of FLOPs of the TC and DU approaches are respectively given by $8M_i\left(2F+1\right)-3F-7$ and $10\left\{M_i\left(\frac{P+1}{2}\right)+2\right\}$ [109], [122]. A summary of the number of FLOPs utilized to asses the computational complexity of various cancelers is shown in Table 5.6.[14]

Based on the aforementioned, in Figs. 5.11a and 5.11b, I depict the FLOPs required by various SI cancelers when tested using the first and second datasets, respectively. From the figures, one can observe that the DU- and TC-based cancelers require the lowest number of FLOPs for all dataset sizes in the first and second datasets, respectively. Further, the polynomial-, RV-TDNN-, and DN-2HLNN-based cancelers require a reasonable number of FLOPs when compared to the others for all dataset

[14]In Table 5.6, I assume for simplicity that each RV and CV activation function costs one and two RV additions, respectively [97].

sizes. Finally, it can be inferred from the figures that the SVR-based cancelers, i.e., RTDSVR and OF-TDSVR require an intolerable computational complexity compared to the others, as their FLOPs depend on the number of support vectors, $N_{sv}^{\Re}$ and $N_{sv}^{\Im}$, as well as the number of testing samples Q [108].

5.6.4.6 Canceler Efficiency

In the previous subsections, I evaluated the performance of each SI canceler in terms of its SIC (or PSD), training overhead, memory storage, and computational complexity. Based on this analysis, I have found that some of the cancelers outperform in terms of SIC performance, and some are promising in terms of training time, memory storage, and/or computational complexity. So, the question is *how to select a certain ML-based SIC approach to fit a target application, i.e., meet system criteria.* This subsection will help to address the above question to select a suitable SIC approach depending on the system requirements.

As the challenge in the SIC problem is to find an SI canceler that maximizes the achieved SIC while minimizing the training time, memory storage, and computational complexity requirements, I have devised an efficiency measure η to evaluate each canceler based on the aforementioned metrics as follows:

$$\eta = \frac{w_{\mathcal{C}}\eta_{\mathcal{C}} + w_{\tau}\eta_{\tau} + w_{\varrho}\eta_{\varrho} + w_{\mathcal{F}}\eta_{\mathcal{F}}}{w_{\mathcal{C}} + w_{\tau} + w_{\varrho} + w_{\mathcal{F}}}, \tag{5.15}$$

where $w_{\mathcal{C}} \in \{0,1\}$, $w_{\tau} \in \{0,1\}$, $w_{\varrho} \in \{0,1\}$, $w_{\mathcal{F}} \in \{0,1\}$ represent the cancellation, training, storage, and complexity weighting factors, respectively, which take either 0 or 1 values depending on the system requirements.[15] Moreover, $\eta_{\mathcal{C}}$, η_{τ}, η_{ϱ}, and $\eta_{\mathcal{F}}$ indicate the cancellation, training, storage, and complexity efficiencies of each

[15]In the following results, I will fix $w_{\mathcal{C}} = 1$ for all test cases as the SIC is the main requirement for any SI canceler.

canceler, which can be respectively expressed as

$$\eta_{\mathcal{C}} = \frac{\mathcal{C} - \mathcal{C}_{min}}{\mathcal{C}_{max} - \mathcal{C}_{min}}, \tag{5.16a}$$

$$\eta_{\tau} = 1 - \frac{\tau - \tau_{min}}{\tau_{max} - \tau_{min}}, \tag{5.16b}$$

$$\eta_{\varrho} = 1 - \frac{\varrho - \varrho_{min}}{\varrho_{max} - \varrho_{min}}, \tag{5.16c}$$

$$\eta_{\mathcal{F}} = 1 - \frac{\mathcal{F} - \mathcal{F}_{min}}{\mathcal{F}_{max} - \mathcal{F}_{min}}, \tag{5.16d}$$

with $\mathcal{C}$ as the total SIC achieved by each canceler over a certain dataset, while $\mathcal{C}_{max}$ and $\mathcal{C}_{min}$ are the maximum and minimum SIC attained by any of the cancelers within this dataset, respectively. Similarly, τ is the training time needed by each canceler over a certain dataset, whereas τ_{max} and τ_{min} are the maximum and minimum training time required by any of the cancelers within this dataset, respectively. Likewise, ϱ represents the number of parameters required by each of the cancelers over a certain dataset, while ϱ_{max} and ϱ_{min} indicate the maximum and minimum parameters needed by any of the cancelers within this dataset, respectively. Finally, $\mathcal{F}$ represents the number of FLOPs required by each of the cancelers over a certain dataset, whereas $\mathcal{F}_{max}$ and $\mathcal{F}_{min}$ denote the maximum and minimum number of FLOPs required by any of the cancelers within this dataset, respectively.

Based on the above, I have assessed the efficiency η for various SI cancelers over the first and second datasets in Table 5.7. It can be observed from the table that the polynomial model achieves the highest efficiency among the other SI cancelers for most of the test cases in the first dataset; i.e., the polynomial-based canceler is efficient for the test cases where a low average transmit power is utilized, and the non-linearity is not severe. However, in the second dataset, where a high transmit power is used, the RV-TDNN-based canceler achieves the highest efficiency among

TABLE 5.7: Efficiency η of different ML-based SI cancelers compared to the polynomial canceler for the first and second datasets.

Dataset	Data size	w_c	w_τ	w_ϱ	$w_\mathcal{F}$	Test case	η									
							Poly.	RV-TDNN	RNN	CV-TDNN	2HLNN	DN-2HLNN	RTDSVR	OF-TDSVR	TC	DU
First dataset (implies lower transmit power, lower non-linear components, and narrower bandwidth)	2000	1	0	0	0	SIC is the only system criterion.		$\mathbf{1^{st}}$		3^{rd}	2^{nd}					
	3000						$\mathbf{1^{st}}$				2^{nd}	3^{rd}				
	4000						$\mathbf{1^{st}}$				2^{nd}	3^{rd}				
	5000						$\mathbf{1^{st}}$			2^{nd}	3^{rd}					
	2000	1	1	0	0	SIC and training time are the only system criteria.		$\mathbf{1^{st}}$		2^{nd}	3^{rd}					
	3000						$\mathbf{1^{st}}$				2^{nd}	3^{rd}				
	4000						$\mathbf{1^{st}}$				2^{nd}	3^{rd}				
	5000						$\mathbf{1^{st}}$			3^{rd}	2^{nd}					
	2000	1	0	1	0	SIC and memory are the only system criteria.					2^{nd}	3^{rd}				1^{st}
	3000						$\mathbf{1^{st}}$				2^{nd}					3^{rd}
	4000						$\mathbf{1^{st}}$				2^{nd}	3^{rd}				
	5000						$\mathbf{1^{st}}$			2^{nd}		3^{rd}				
	2000	1	0	0	1	SIC and complexity are the only system criteria.		$\mathbf{1^{st}}$		2^{nd}	3^{rd}					
	3000						$\mathbf{1^{st}}$				2^{nd}	3^{rd}				
	4000						$\mathbf{1^{st}}$				2^{nd}	3^{rd}				
	5000						$\mathbf{1^{st}}$			2^{nd}	3^{rd}					
	2000	1	1	1	0	SIC, training time, and memory are the only system criteria.					$\mathbf{1^{st}}$	3^{rd}				2^{nd}
	3000						$\mathbf{1^{st}}$				2^{nd}	3^{rd}				
	4000						$\mathbf{1^{st}}$				2^{nd}	3^{rd}				
	5000						$\mathbf{1^{st}}$				3^{rd}	2^{nd}				
	2000	1	1	0	1	SIC, training time, and complexity are the only system criteria.		$\mathbf{1^{st}}$		2^{nd}	3^{rd}					
	3000						$\mathbf{1^{st}}$				2^{nd}	3^{rd}				
	4000						$\mathbf{1^{st}}$				2^{nd}	3^{rd}				
	5000						$\mathbf{1^{st}}$				3^{rd}	2^{nd}				
	2000	1	0	1	1	SIC, memory, and complexity are the only system criteria.					2^{nd}	3^{rd}				$\mathbf{1^{st}}$
	3000						$\mathbf{1^{st}}$				2^{nd}					3^{rd}
	4000						$\mathbf{1^{st}}$				2^{nd}	3^{rd}				
	5000						$\mathbf{1^{st}}$			2^{nd}	3^{rd}					
	2000	1	1	1	1	SIC, training time, memory, and complexity are all system criteria.					2^{nd}	3^{rd}				$\mathbf{1^{st}}$
	3000						$\mathbf{1^{st}}$				2^{nd}	3^{rd}				
	4000						$\mathbf{1^{st}}$				2^{nd}	3^{rd}				
	5000						$\mathbf{1^{st}}$				3^{rd}	2^{nd}				
Second dataset (implies higher transmit power, higher non-linear components, and wider bandwidth)	2000	1	0	0	0	SIC is the only system criterion.		$\mathbf{1^{st}}$	3^{rd}	2^{nd}						
	3000							$\mathbf{1^{st}}$			3^{rd}	2^{nd}				
	4000						3^{rd}	$\mathbf{1^{st}}$	2^{nd}							
	5000							$\mathbf{1^{st}}$	2^{nd}		3^{rd}					
	2000	1	1	0	0	SIC and training time are the only system criteria.	2^{nd}	3^{rd}						$\mathbf{1^{st}}$		
	3000						$\mathbf{1^{st}}$		2^{nd}		3^{rd}					
	4000						$\mathbf{1^{st}}$	2^{nd}	3^{rd}							
	5000						3^{rd}	$\mathbf{1^{st}}$	2^{nd}							
	2000	1	0	1	0	SIC and memory are the only system criteria.		$\mathbf{1^{st}}$		2^{nd}						3^{rd}
	3000							$\mathbf{1^{st}}$			3^{rd}	2^{nd}				
	4000						2^{nd}	$\mathbf{1^{st}}$								3^{rd}
	5000						2^{nd}	$\mathbf{1^{st}}$			3^{rd}					
	2000	1	0	0	1	SIC and complexity are the only system criteria.		$\mathbf{1^{st}}$	3^{rd}	2^{nd}						
	3000							$\mathbf{1^{st}}$			3^{rd}	2^{nd}				
	4000						3^{rd}	$\mathbf{1^{st}}$	2^{nd}							
	5000							$\mathbf{1^{st}}$	2^{nd}		3^{rd}					
	2000	1	1	1	0	SIC, training time, and memory are the only system criteria.	$\mathbf{1^{st}}$	2^{nd}			3^{rd}					
	3000						$\mathbf{1^{st}}$				2^{nd}	3^{rd}				
	4000						$\mathbf{1^{st}}$	2^{nd}				3^{rd}				
	5000						2^{nd}	$\mathbf{1^{st}}$				3^{rd}				
	2000	1	1	0	1	SIC, training time, and complexity are the only system criteria.	$\mathbf{1^{st}}$	2^{nd}			3^{rd}					
	3000						$\mathbf{1^{st}}$			2^{nd}	3^{rd}					
	4000						$\mathbf{1^{st}}$	2^{nd}	3^{rd}							
	5000						3^{rd}	$\mathbf{1^{st}}$	2^{nd}							
	2000	1	0	1	1	SIC, memory, and complexity are the only system criteria.		$\mathbf{1^{st}}$		2^{nd}	3^{rd}					
	3000							$\mathbf{1^{st}}$			3^{rd}	2^{nd}				
	4000						2^{nd}	$\mathbf{1^{st}}$								3^{rd}
	5000						2^{nd}	$\mathbf{1^{st}}$				3^{rd}				
	2000	1	1	1	1	SIC, training time, memory, and complexity are all system criteria.	$\mathbf{1^{st}}$	2^{nd}				3^{rd}				
	3000						$\mathbf{1^{st}}$				2^{nd}	3^{rd}				
	4000						$\mathbf{1^{st}}$	2^{nd}			3^{rd}					
	5000						2^{nd}	$\mathbf{1^{st}}$			3^{rd}					

The best three cancelers that achieve the highest efficiency are ranked by the 1^{st}, 2^{nd}, and 3^{rd}, respectively.

the others for most of the test cases. One can also notice from Table 5.7 that the polynomial-based canceler requires a large number of training examples to achieve the highest efficiency, e.g., the polynomial-based canceler is unable to attain the highest efficiency when being trained using 2000 samples of the first dataset. In addition, one can infer from the table that the RV-TDNN works well in the test cases where the training overhead is not of the system demands, e.g., the RV-TDNN-based canceler is unable to attain the highest efficiency in the second dataset for all test cases where $w_\tau = 1$ and the polynomial-based canceler becomes a better choice in such test cases.

In sum, upon testing several ML-based approaches for SIC in FD transceivers, using two test setups and over short dataset sizes, I can conclude that the model-driven approaches, i.e., polynomial-based canceler, can be a good choice in operating scenarios where a low transmit power is employed; however, at high transmit power levels, the data-driven ML approaches, i.e., RV-TDNN-based canceler, can be a better choice.

5.7 Challenges and Future Research Directions

The previous sections provided a comprehensive overview of applying ML-based approaches for SIC in FD transceivers. Suitable SIC approaches have also been selected for SIC, depending on the system criteria. Although the literature works surveyed in this chapter provide a significant role in empowering the application of ML techniques for SIC in FD transceivers, more efforts remain to be made to adopt such techniques in practical wireless systems employing FD transmission. The following subsections delve into the main challenges of applying ML-based approaches for SIC in FD transceivers and provide a guide for future research directions.

5.7.1 Considering the Effect of SoI while Performing the SIC

The existing ML-based SIC approaches consider the cancellation of the SI signal only, i.e., no signal from any remote FD or half-duplex TRPs is considered. However, in practical situations, i.e., real-time FD systems, the SIC in one FD node has to be done while an SoI from another TRP is received and demodulated. Initial works in [127], [130] investigated a joint detection of the SI and SoI and proved that an NN-based SI canceler is beneficial to enhance the signal demodulation. Despite the potential of the works in [127], [130], there are still more issues remaining to be addressed, and the point of detecting the SoI while performing the SIC is open to improvements from both performance and complexity perspectives. For instance, one issue is that all ML-based approaches surveyed in this chapter are trained and verified using time-domain samples, i.e., they are completely working in the time domain. However, if the SoI signal employs any of the frequency-domain modulation formats, e.g., OFDM modulation, performing the SIC could be done in the frequency domain; this would be similar to the fifth-generation new radio or future 6G demodulation pilots (demodulation reference signals uplink or downlink) which are in specific time and frequency symbols [127]. Thus, adapting the ML-based SIC approaches to work with frequency- rather than time-domain samples can be a direction for future investigation.

5.7.2 Tackling the Time-Varying SI Channels

The existing ML-based SIC approaches use offline-trained ML algorithms to estimate the SI signal over a static SI channel. However, in practical situations, the movements of user equipment TRPs and/or environmental changes can vary the SI channel over time, and the ML algorithms may need to be retrained in order to adapt to the time-varying SI channel. Nevertheless, as presented in Fig. 5.9, some ML algorithms

require a higher training time, i.e., they are not fast enough to be retrained during the FD transmission, which can lead to significant performance degradation. Initial works in [102], [131] investigate the effect of canceling the SI signal under time-varying SI channels. However, these are incipient works, and the point is open to improvements in both performance and complexity perspectives. For instance, applying reinforcement and online learning to iteratively tackle the time-varying SI channel can be a future direction of investigation. Scaling the performance and/or complexity as a result of employing reinforcement and online learning can also be considered for future investigation.

5.7.3 Applying ML Approaches for SIC in FD MIMO Systems

The ML-based SIC approaches surveyed in this work are trained and verified using a single-input single-output (SISO) FD testbed. However, in recent communication standards, the MIMO technology has become a basic transmit/receive scheme. Hence, extending the above ML-based SIC techniques to MIMO rather than SISO FD transceivers is imperative. Typically, the complexity of the SIC approaches exponentially increases under MIMO operation where M transmit antennas interfere with N receive antennas. A straightforward approach—to process several SI signals in the digital domain—is to perform the SIC using separate SI cancelers, which consider the interfering signals from all transmit antennas; however, this results in excessive complexity. To address this issue, alternative approaches can be designed. For instance, exploiting the spatial correlation between the MIMO channels to develop a common SI canceler, i.e., not separate cancelers, can be a direction for future investigation in order to reduce the impractical computational complexity of the traditional MIMO SIC-based approaches [40].

5.7.4 Training Complexity of ML-based SIC Approaches

The computational complexity of the existing ML-based SIC approaches is typically evaluated and compared in terms of FLOPs required in the inference stage, i.e., upon performing and finalizing the training process. However, estimating the training complexity (in terms of FLOPs) is crucial and should be considered, especially for ML algorithms targeted to be integrated with online learning as described in Section 5.7.2. For instance, calculating the number of FLOPs required for performing the backpropagation in NNs, approximating the unknown function using optimization in SVRs, and solving the low-rank tensor decomposition problem in TC-based cancelers should be explored to provide insights about the feasibility of applying ML-based approaches for SIC in real-time FD transceivers.

5.8 Conclusions

In this chapter, I have surveyed the up-to-date contributions in applying ML approaches for SIC in FD transceivers. Based on a comprehensive review, I have found that canceling the interference in FD transceivers using ML has been initially performed by traditional approaches, such as NNs and SVRs. Advanced ML approaches, such as TC, TensorFlow graphs, and RFFs, integrated with online learning, have been employed for SIC as well. Further, other ML approaches proven in other disciplines, such as DR, GMMs, DU, LL, and APSM, have also been utilized for modeling the SI in FD transceivers. Upon surveying the literature, I have provided a case study to evaluate the performance of the prominent ML-based approaches over short dataset sizes and using two test setups employing different transmit power levels. Specifically, I have assessed the performance of the prominent data-driven ML-based approaches in terms of the SIC, PSD, training time, memory storage, and compu-

tational complexity and compared them with those of the model-driven approaches, e.g., polynomial-based canceler. Afterward, I evaluated the efficiency of the different SIC approaches based on the aforementioned metrics to select a suitable approach for SIC, depending on system requirements. Based on this study, I have found that the model-driven approaches, i.e., polynomial-based canceler, could be a good choice when a low transmit power is utilized (i.e., low non-linearity exists). However, at high transmit power (i.e., high non-linearity exists), the data-driven ML-based approaches, i.e., RV-TDNN-based canceler, could be a better choice. I have finally identified the research gaps in applying ML approaches for SIC in FD transceivers, paving the way for future research directions, such as considering the SoI effect, extension to MIMO FD transceivers, and tackling the time-varying SI channels.

References

[1] L. Bariah, L. Mohjazi, S. Muhaidat, P. C. Sofotasios, G. K. Kurt, H. Yanikomeroglu, and O. A. Dobre, "A prospective look: Key enabling technologies, applications and open research topics in 6G networks," *IEEE Access*, vol. 8, pp. 174792–174820, Aug. 2020.

[2] P. Yang, Y. Xiao, M. Xiao, and S. Li, "6G wireless communications: Vision and potential techniques," *IEEE Netw.*, vol. 33, no. 4, pp. 70–75, Jul. 2019.

[3] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?," *Nature Electron.*, vol. 3, no. 1, pp. 20–29, Jan. 2020.

[4] X. You *et al.*, "Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Sci. China Inf. Sci.*, vol. 64, no. 1, pp. 1–74, Nov. 2020.

[5] Z. Zhang *et al.*, "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 28–41, Sep. 2019.

[6] E. C. Strinati *et al.*, "6G: The next frontier: From holographic messaging to artificial intelligence using subterahertz and visible light communication," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 42–50, Sep. 2019.

[7] K. David and H. Berndt, "6G vision and requirements: Is there any need for beyond 5G?" *IEEE Veh. Technol. Mag.*, vol. 13, no. 3, pp. 72–80, Sep. 2018.

[8] S. Zhang, C. Xiang, and S. Xu, "6G: Connecting everything by 1000 times price reduction," *IEEE Open J. Veh. Technol.*, vol. 1, pp. 107–115, Mar. 2020.

[9] H. Viswanathan and P. E. Mogensen, "Communications in the 6G era," *IEEE Access*, vol. 8, pp. 57063–57074, Mar. 2020.

[10] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, Mar. 2020.

[11] S. Chen, Y.-C. Liang, S. Sun, S. Kang, W. Cheng, and M. Peng, "Vision, requirements, and technology trend of 6G: How to tackle the challenges of system coverage, capacity, user data-rate and movement speed," *IEEE Wireless Commun. Mag.*, vol. 27, no. 2, pp. 218–228, Apr. 2020.

[12] F. Tariq, M. R. A. Khandaker, K.-K. Wong, M. A. Imran, M. Bennis, and M. Debbah, "A speculative study on 6G," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 118–125, Aug. 2020.

[13] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 957–975, Jul. 2020.

[14] I. F. Akyildiz, A. Kak, and S. Nie, "6G and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133995–134030, July 2020.

[15] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May 2020.

[16] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, Aug. 2019.

[17] B. Zong, C. Fan, X. Wang, X. Duan, B. Wang, and J. Wang, "6G technologies: Key drivers, core requirements, system architectures, and enabling technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 18–27, Sep. 2019.

[18] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 334–366, Feb. 2021.

[19] A. Shahraki, M. Abbasi, M. J. Piran, and A. Taherkordi, "A comprehensive survey on 6G networks: Applications, core services, enabling technologies, and future challenges," Jan. 2021. [Online]. Available: arXiv:2101.12475.

[20] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, D. Niyato, O. A. Dobre, H. V. Poor, "6G internet of things: A comprehensive survey," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 359–383, Jan. 2022.

[21] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

[22] H. V. Nguyen, V.-D. Nguyen, O. A. Dobre, and O.-S. Shin, "Sum rate maximization based on sub-array antenna selection in a full-duplex system," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.

[23] A. Yadav, O. A. Dobre, and N. Ansari, "Energy and traffic aware full-duplex communications for 5G systems," *IEEE Access*, vol. 5, pp. 11278–11290, May 2017.

[24] A. Yadav, O. A. Dobre, and H. V. Poor, "Is self-interference in full-duplex communications a foe or a friend?" *IEEE Signal Process. Lett.*, vol. 25, no. 7, pp. 951–955, Jul. 2018.

[25] E. Everett, A. Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 680–694, Jan. 2014.

[26] S. Hong *et al.*, "Applications of self-interference cancellation in 5G and beyond," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 114–121, Feb. 2014.

[27] D. Korpi *et al.*, "Full-duplex transceiver system calculations: Analysis of ADC and linearity challenges," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3821–3836, Jul. 2014.

[28] 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.

[29] A. Yadav, G. I. Tsiropoulos, and O. A. Dobre, "Full-duplex communications: Performance in ultradense mm-wave small-cell wireless networks," *IEEE Veh. Technol. Mag.*, vol. 13, no. 2, pp. 40–47, Jun. 2018.

[30] Z. Zhang, X. Chai, K. Long, A. V. Vasilakos, and L. Hanzo, "Full duplex techniques for 5G networks: Self-interference cancellation, protocol design, and relay selection," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 128–137, May 2015.

[31] M. Chung, M. S. Sim, J. Kim, D. K. Kim, and C.-B. Chae, "Prototyping real-time full duplex radios," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 56–63, Sep. 2015.

[32] D. Kim, H. Lee, and D. Hong, "A survey of in-band full-duplex transmission: From the perspective of PHY and MAC layers," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2017–2046, 4th quart., 2015.

[33] Z. Zhang, K. Long, A. V. Vasilakos, and L. Hanzo, "Full-duplex wireless communications: Challenges, solutions, and future research directions," *Proc. IEEE*, vol. 104, no. 7, pp. 1369–1409, Jul. 2016.

[34] D. Korpi *et al.*, "Full-duplex mobile device: Pushing the limits," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 80–87, Sep. 2016.

[35] M. Amjad, F. Akhtar, M. H. Rehmani, M. Reisslein, and T. Umer, "Full-duplex communication in cognitive radio networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2158–2191, 4th Quart., 2017.

[36] V.-D. Nguyen, H. V. Nguyen, O. A. Dobre, and O.-S. Shin, "A new design paradigm for secure full-duplex multiuser systems," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 7, pp. 1480–1498, Jul. 2018.

[37] H. V. Nguyen, V.-D. Nguyen, O. A. Dobre, Y. Wu, and O.-S. Shin, "Joint antenna array mode selection and user assignment for full-duplex MU-MISO systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 2946–2963, Jun. 2019.

[38] H. V. Nguyen, V.-D. Nguyen, O. A. Dobre, D. N. Nguyen, E. Dutkiewicz, and O.-S. Shin, "Joint power control and user association for NOMA-based full-

duplex systems," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 8037–8055, Nov. 2019.

[39] Q. N. Le, N.-P. Nguyen, A. Yadav, and O. A. Dobre, "Outage performance of full-duplex overlay CR-NOMA networks with SWIPT," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[40] H. V. Nguyen, V.-D. Nguyen, O. A. Dobre, N.-P. Nguyen, R. Zhao, and S. Chatzinotas, "On the spectral and energy efficiencies of full-duplex cell-free massive MIMO," *IEEE J. Select. Areas Commun.*, vol. 38, no. 8, pp. 1698–1718, Aug. 2020.

[41] Q. N. Le, A. Yadav, N.-P. Nguyen, O. A. Dobre, and R. Zhao, "Full-duplex non-orthogonal access cooperative overlay spectrum-sharing networks with SWIPT," *IEEE Trans. Green Commun. Netw.*, vol. 5, no. 1, pp. 322–334, Mar. 2021.

[42] E. A. Makled and O. A. Dobre, "On the security of full-duplex relay-assisted underwater acoustic network with NOMA," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6255–6265, Jun. 2022.

[43] 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.

[44] B. Debaillie *et al.*, "Analog/RF solutions enabling compact full-duplex radios," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1662–1673, Sep. 2014.

[45] Y.-S. Choi and H. Shirani-Mehr, "Simultaneous transmission and reception: Algorithm, design and system level performance," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 12, pp. 5992–6010, Dec. 2013.

[46] L. Laughlin, M. A. Beach, K. A. Morris, and J. L. Haine, "Optimum single antenna full duplex using hybrid junctions," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1653–1661, Sep. 2014.

[47] D. Korpi, L. Anttila, V. Syrjala, and M. Valkama, "Widely linear digital self-interference cancellation in direct-conversion fullduplex transceiver," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1674–1687, Sep. 2014.

[48] D. Korpi, L. Anttila, and M. Valkama, "Reference receiver based digital self-interference cancellation in MIMO full-duplex transceivers," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 1001–1007.

[49] D. Korpi, *et al*, "Digital self-interference cancellation under nonideal RF components: Advanced algorithms and measured performance," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2015, pp. 286–290.

[50] E. Ahmed and A. M. Eltawil, "All-digital self-interference cancellation technique for full-duplex systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3519–3532, Jul. 2015.

[51] D. Korpi, Y.-S. Choi, T. Huusari, L. Anttila, S. Talwar, and M. Valkama, "Adaptive nonlinear digital self-interference cancellation for mobile inband full-duplex radio: Algorithms and RF measurements," in *Proc. IEEE IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–7.

[52] M. A. Tafreshi, M. Koskela, D. Korpi, P. Jääskeläinen, M. Valkama, and J. Takala, "Software defined radio implementation of adaptive nonlinear digital self-interference cancellation for mobile inband fullduplex radio," in *Proc. IEEE Global Conf. Signal Inform. Process. (GLOBALSIP)*, Dec. 2016, pp. 733–737.

[53] M. S. Amjad and O. Gurbuz, "Linear digital cancellation with reduced computational complexity for full-duplex radios," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[54] X. Quan, Y. Liu, S. Shao, C. Huang, and Y. Tang, "Impacts of phase noise on digital self-interference cancellation in full-duplex communications," *IEEE Trans. Signal Process.*, vol. 65, no. 7, pp. 1881–1893, Apr. 2017.

[55] Y. Liu, X. Quan, W. Pan, and Y. Tang, "Digitally assisted analog interference cancellation for in-band full-duplex radios," *IEEE Commun. Lett.*, vol. 21, no. 5, pp. 1079–1082, May 2017.

[56] P. Ferrand and M. Duarte, "Multi-tap digital canceller for full-duplex applications," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2017, pp. 1–5.

[57] D. Korpi, L. Anttila, and M. Valkama, "Nonlinear self-interference cancellation in MIMO full-duplex transceivers under crosstalk," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, pp. 1–15, Dec. 2017.

[58] R. V. Kulkarni, A. Forster, and G. K. Venayagamoorthy, "Computational intelligence in wireless sensor networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 1, pp. 68–96, 1st Quart., 2011.

[59] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine-learning techniques in cognitive radios," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1136–1159, 3rd Quart., 2013.

[60] M. A. Alsheikh, S. Lin, D. Niyato, and H.-P. Tan, "Machine learning in wireless sensor networks: Algorithms, strategies, and applications," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1996–2018, 4th Quart., 2014.

[61] H. A. Al-Rawi, M. A. Ng, and K.-L. A. Yau, "Application of reinforcement learning to routing in distributed wireless networks: A review," *Artif. Intell. Rev.*, vol. 43, no. 3, pp. 381–416, Jan. 2015.

[62] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A survey of machine learning techniques applied to self-organizing cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2392–2431, 4th Quart., 2017.

[63] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[64] E. Nachmani, E. Marciano, L. Lugosch, W. J. Gross, D. Burshtein, and Y. Be'ery, "Deep learning methods for improved decoding of linear codes," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 119–131, Feb. 2018.

[65] F. Liang, C. Shen, and F. Wu, "An iterative BP-CNN architecture for channel decoding," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 144–159, Feb. 2018.

[66] F. Pacheco, E. Exposito, M. Gineste, C. Baudoin, and J. Aguilar, "Towards the deployment of machine learning solutions in network traffic classification: A systematic survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1988–2014, 2nd Quart., 2018.

[67] D. He, C. Liu, T. Q. S. Quek, and H. Wang, "Transmit antenna selection in MIMO wiretap channels: A machine learning approach," *IEEE Wireless Commun. Lett.*, vol. 7, no. 4, pp. 634–637, Aug. 2018.

[68] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 852–855, Oct. 2018.

[69] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2595–2621, 4th Quart., 2018.

[70] M. Gao, Y. Li, O. A. Dobre, and N. Al-Dhahir, "Joint blind identification of the number of transmit antennas and MIMO schemes using Gerschgorin radii and FNN," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 373–387, Jan. 2019.

[71] J. Xie *et al.*, "A survey of machine learning techniques applied to software defined networking (SDN): Research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 393–430, 1st Quart., 2019.

[72] M. Usama *et al.*, "Unsupervised machine learning for networking: Techniques, applications and research challenges," *IEEE Access*, vol. 7, pp. 65579–65615, May 2019.

[73] X. Cheng, D. Liu, C. Wang, S. Yan, and Z. Zhu, "Deep learning-based channel estimation and equalization scheme for fbmc/oqam systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 881–884, Jun. 2019.

[74] M. Chen, *et al.*, "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3039–3071, 4th Quart., 2019.

[75] N. C. Luong *et al.*, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3133–3174, 4th Quart., 2019.

[76] Y. Sun, M. Peng, Y. Zhou, Y. Huang, and S. Mao, "Application of machine learning in wireless networks: Key techniques and open issues," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3072–3108, 4th Quart., 2019.

[77] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation, opportunities and challenges," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 46–51, Jun. 2020.

[78] A. Faisal, I. Al-Nahhal, O. A. Dobre, and T. M. N. Ngatched, "Deep reinforcement learning for optimizing RIS-assisted HD-FD wireless systems," *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3893–3897, Dec. 2021.

[79] S. Zhang, S. Zhang, F. Gao, J. Ma, and O. A. Dobre, "Deep learning based RIS channel extrapolation with element-grouping," *IEEE Wireless Commun. Lett.*, vol. 10, no. 12, pp. 2644–2648, Dec. 2021.

[80] M. Xu, S. Zhang, J. Ma, and O. A. Dobre, "Deep learning-based time-varying channel estimation for RIS assisted communication," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 94–98, Jan. 2022.

[81] A. Faisal, I. Al-Nahhal, O. A. Dobre, and T. M. N. Ngatched, "Deep reinforcement learning for RIS-assisted FD systems: Single or distributed RIS?," *IEEE Commun. Lett.*, vol. 26, no. 7, pp. 1563–1567, Jul. 2022.

[82] M. Al-Nahhal, I. Al-Nahhal, O. A. Dobre, X. Lin, D. Chang, and C. Li, "Joint estimation of linear and nonlinear coherent optical fiber signal to-noise ratio," *IEEE Photon. Technol. Lett.*, vol. 35, no. 1, pp. 23–26, Nov. 2022.

[83] M. Al-Nahhal, I. Al-Nahhal, O. A. Dobre, S. K. O. Soman, D. Chang, and C. Li, "Learned signal-to-noise ratio estimation in optical fiber communication links," *IEEE Photon. J.*, vol. 14, no. 6, pp. 1–7, Nov. 2022.

[84] A. Faisal, I. Al-Nahhal, O. A. Dobre, and T. M. N. Ngatched, "Distributed RIS-assisted FD systems with discrete phase shifts: A reinforcement learning approach," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2022, pp. 5862–5867.

[85] Y. Liu, I. Al-Nahhal, O. A. Dobre, and F. Wang, "Deep-learning-based channel estimation for IRS-assisted ISAC system," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2022, pp. 4220–4225.

[86] Y. Liu, I. Al-Nahhal, O. A. Dobre, and F. Wang, "Deep-learning channel estimation for IRS-assisted integrated sensing and communication system," *IEEE Trans. Veh Technol.*, early access, pp. 1–14, Dec. 2022.

[87] E. A. Makled, I. Al-Nahhal, O. A. Dobre, and O. Üreten, "Identification of cellular signal measurements using machine learning," *IEEE Trans. Instrum. Meas.*, vol. 72, no. 1, pp. 1–4, Jan. 2023.

[88] 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.

[89] 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.

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

[91] 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.*, Dec. 2019, pp. 53–57.

[92] K. E. Kolodziej, A. U. Cookson, and B. T. Perry, "Neural network tuning for analog-RF self-interference cancellation," in *Proc. IEEE MTT-S Int. Microw. Symp.*, June 2021, pp. 673–676.

[93] K. E. Kolodziej, A. U. Cookson, and B. T. Perry, "RF canceller tuning acceleration using neural network machine learning for in-band full-duplex systems," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1330–1334, May 2021.

[94] V. Tapio and M. Juntti, "Non-linear self-interference cancelation for full-duplex transceivers based on Hammerstein-Wiener model," *IEEE Commun. Lett.*, vol. 25, no. 11, pp. 3684–3688, Nov. 2021.

[95] 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. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 589–593.

[96] 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.

[97] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Advanced machine learning techniques for self-interference cancellation in full-duplex radios," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 1149–1153.

[98] Q. Wang, F. He, and J. Meng, "Performance comparison of real and complex valued neural networks for digital self-interference cancellation," in *Proc. IEEE 19th Int. Conf. Commun. Technol. (ICCT)*, Oct. 2019, pp. 1193–1199.

[99] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Low complexity neural network structures for self-interference cancellation in full-duplex radio," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 181–185, Jan. 2021.

[100] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Hybrid-layers neural network architectures for modeling the self-interference in full-duplex systems," *IEEE Trans. Veh Technol.*, vol. 71, no. 6, pp. 6291–6307, Jun. 2022.

[101] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Full-duplex self-interference cancellation using dual-neurons neural networks," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 557–561, Mar. 2022.

[102] D. H. Kong, Y.-S. Kil, and S.-H. Kim, "Neural network aided digital self-interference cancellation for full-duplex communication over time-varying channels," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6201–6213, Jun. 2022.

[103] 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 Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[104] W. Zhang, J. Yin, D. Wu, G. Guo, and Z. Lai, "A self-interference cancellation method based on deep learning for beyond 5G full-duplex system," in *Proc.*

IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC), Sep. 2018, pp. 1–5.

[105] K. Muranov, M. A. Islam, B. Smida, and N. Devroye, "On deep learning assisted self-interference estimation in a full-duplex relay link," *IEEE Wireless Commun. Lett.*, vol. 10, no. 12, pp. 2762–2766, Dec. 2021.

[106] C. Auer, K. Kostoglou, T. Paireder, O. Ploder, and M. Huemer, "Support vector machines for self-interference cancellation in mobile communication transceivers," in *Proc. IEEE Veh. Technol. Conf. (VTC-Spring)*, Jun. 2020, pp. 1–6.

[107] M. Erdem, H. Ozkan, and O. Gurbuz, "Nonlinear digital self-interference cancellation with SVR for full duplex communication," in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC)*, Jun. 2020, pp. 1–6.

[108] M. Yilan, O. Gurbuz, and H. Ozkan, "Integrated linear and nonlinear digital cancellation for full duplex communication," *IEEE Wireless Communications*, vol. 28, no. 1, pp. 20–27, Feb. 2021.

[109] F. Jochems and A. Balatsoukas-Stimming, "Non-linear self-interference cancellation via tensor completion," in *Proc. 54th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2020, pp. 905–909.

[110] 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. (ICNC)*, Mar. 2018, pp. 779–783.

[111] K. Wongsuphasawat, D. Smilkov, J. Wexler, J. Wilson, D. Mane, D. Fritz, D. Krishnan, F. B. Viegas, and M. Wattenberg, "Visualizing dataflow graphs of

deep learning models in TensorFlow," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 1, pp. 1–12, Jan. 2018.

[112] M. Erdem, H. Ozkan, and O. Gurbuz, "A new online nonlinear self-interference cancelation method with random Fourier features," *IEEE Wireless Commun. Lett.*, vol. 11, no. 7, pp. 1379–1383, Apr. 2022.

[113] R. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Proc. 20th Int. Conf. Adv. Neural Inf. Process. Syst.*, Dec. 2007, pp. 1177–1184.

[114] J. Lu, S. C. Hoi, J. Wang, P. Zhao, and Z.-Y. Liu, "Large scale online kernel learning," *J. Mach. Learn. Res.*, vol. 17, no. 47, pp. 1–43, Jan. 2016.

[115] S. Mehrkanoon and J. A. K. Suykens, "Deep hybrid neural-kernel networks using random Fourier features," *Neurocomputing*, vol. 298, no. 7, pp. 46–54, Jul. 2018.

[116] R. Mitra, V. Bhatia, S. Jain, and K. Choi, "Performance analysis of random Fourier features based unsupervised multistage-clustering for VLC," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2659–2663, Aug. 2021.

[117] R. Mitra, *et al.*, "Analytical guarantees for hyperparameter free RFF based deep learning in the low-data regime," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 2, pp. 634–638, Feb. 2022.

[118] R. Mitra and G. Kaddoum, "Random Fourier feature based deep learning for wireless communications," *IEEE Trans. Cogn. Commun. Netw.*, vol. 11, no. 2, pp. 468–479., Jun. 2022.

[119] Y.-H. Lin, Y.-T. Liao, J.-Y. Chu, P.-J. Su, and T.-Y. Hsu, "Digital self interference cancellation via dynamic regression for in-band full-duplex system," in *Proc. IEEE 5th Global Conf. Consum. Electron.*, Oct. 2016, pp. 1–3.

[120] J. Chen, L. Zhang, and Y.-C. Liang, "Exploiting Gaussian mixture model clustering for full-duplex transceiver design," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5802–5816, Aug. 2019.

[121] X. F. He, D. Cai, Y. L. Shao, H. J. Bao, and J. W. Han, "Laplacian regularized Gaussian mixture model for data clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 9, pp. 1406–1418, Sep. 2011.

[122] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Identification of nonlinear RF systems using backpropagation," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2020, pp. 1–6.

[123] J. R. Hershey, J. Le Roux, and F. Weninger, "Deep unfolding: Model-based inspiration of novel deep architectures," Nov. 2014, arXiv:1409.2574. [Online]. Available: http://arxiv.org/abs/1409.2574.

[124] A. Balatsoukas-Stimming and C. Studer, "Deep unfolding for communications systems: A survey and some new directions," in *Proc. IEEE Int. Workshop Signal Process. Syst. (SiPS)*, Oct. 2019, pp. 266–271.

[125] O. Zhao, W-S Liao, K. Li, T. Matsumura, F. Kojima, and H. Harada, "Lazy learning-based self-interference cancellation for wireless communication systems with in-band full-duplex operations," in *Proc. IEEE 32nd Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2021, pp. 1589–1594.

[126] M. H. Attar *et* al., "Parallel APSM for fast and adaptive digital SIC in full-duplex transceivers with nonlinearity," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2022, pp. 1–5.

[127] A. Balatsoukas-Stimming, "Joint detection and self-interference cancellation in full-duplex systems using machine learning," in *Proc. 55th Asilomar Conf. Signals, Syst., Comput.*, Mar. 2021, pp. 989–992.

[128] T. Liu *et al.*, "Dynamic behavioral modeling of 3G power amplifiers using real-valued time delay neural networks," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 3, pp. 1025–1033, Mar. 2004.

[129] J. Cai, C. Yu, L. Sun, S. Chen, and J. B. King, "Dynamic behavioral modeling of RF power amplifier based on time-delay support vector regression," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 2, pp. 533–543, Feb. 2019.

[130] A. Balatsoukas-Stimming, "End-to-end learned self-interference cancellation," in *Proc. 56th Asilomar Conf. Signals, Syst., Comput.*, Mar. 2022, pp. 1334–1338.

[131] O. Ploder *et al.*, "SICNet—Low complexity sample adaptive neural network-based self-interference cancellation in LTE-A/5G mobile transceivers," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 958–972, Jun. 2022.

[132] Y. Chen *et. al.*, "MIMO full duplex radios with deep learning," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jul. 2020, pp. 1–6.

Chapter 6

Residual Neural Networks for Learning the Full-Duplex Self-Interference

6.1 Abstract

Full-duplex (FD) is a key technology for enhancing the capacity of next-generation wireless systems by jointly maximizing the utilization of time and frequency resources, resulting in low latency and high spectral efficiency. However, the self-interference (SI), leaking to the receiver chain from its own transmitter chain, is the main issue that hinders reaping the key benefits of FD systems, and SI cancellation (SIC) is introduced to enable such benefits. Digital non-linear SIC is traditionally performed using model-driven approaches, such as polynomial models, which are of high complexity. Thus, data-driven machine learning (ML) approaches are introduced for learning the FD non-linear SI with lower complexity. This chapter proposes an ML approach based on residual neural network (Res-NN) to learn the FD non-linear SI and relax

the computational requirements of the traditional methods. Res-NN uses shortcut connections from the input/hidden layer to the output layer to enhance the learning capabilities of the SI cancelers. Simulation results show that an NN employing residual connections could effectively learn the FD SI and outperform the existing benchmarks in the literature.

6.2 Introduction

Full-duplex (FD), maximizing the utilization of time and frequency resources, has become one of the emerging technologies for widespread applications ranging from wireless and vehicular to optical and underwater communications [1]-[4]. However, the self-interference (SI)—a loop-back signal from the transmitter (Tx) chain to its co-located receiver (Rx) chain—is the main issue that hinders reaping the key benefits of FD systems [5]. Such interference is crucial to the performance of FD systems as it significantly degrades their ability to detect the desired signal of interest (SoI). Reaping the benefits of FD systems can only be realized by applying various self-interference cancellation (SIC) techniques, which can be done in three domains: *propagation, analog, and/or digital* [5], [6].

Propagation domain cancellation can be performed around the antennas using passive or active elements, such as antenna separation, phase control, surface treatment, coupling networks, duplexers, and/or circulators [5], [6]. Analog domain cancellation can be done in the analog circuity after performing the propagation domain cancellation and before applying the analog-to-digital converter (ADC). Analog domain cancellation generates a copy of the SI signal using techniques such as auxiliary Tx chains [5], [6]. Digital domain cancellation is performed after applying the ADC using model-driven approaches such as linear, widely-linear, and polynomial models, e.g.,

Wiener, Hammerstein, and parallel Hammerstein [5], [6]. Digital domain cancellation using polynomial models is shown to be effective for non-linear SIC in FD transceivers. However, they suffer from a higher computational complexity [7]. Thus, data-driven machine learning (ML) approaches have been recently introduced for learning the FD non-linear SI with reduced complexity.

For the time being, there have been significant advances in applying ML approaches for SIC in FD transceivers. Approaches such as neural networks (NNs) [8]-[13], support vector regressors (SVRs) [14], [15], tensor completion [16], deep unfolding [17], and others [18],[1] have been introduced to learn the SI with lower complexity than the traditional polynomial model. The synergy between ML and FD communications was shown to be efficient in terms of computational resources when compared to the polynomial model. However, further reduction in these resources is required to enable the synergy of ML with real-time FD transmit and receive points (TRPs). This chapter fills this gap by proposing an ML solution based on residual-NN (Res-NN) in order to learn the SI with reduced computational resources. The proposed Res-NN exploits shortcut connections from the hidden to the output layer to enhance the learning capabilities of the SI cancelers. The results show that an NN employing residual connections could effectively learn the FD non-linear SI and outperform the existing benchmarks in the literature. Digital SIC, power spectral density (PSD), network parameters, and computational complexity evaluations are provided to support the findings.

The rest of this chapter is structured as follows. The system model integrating Res-NN with FD communications is introduced in Section 6.3. The basic concepts of residual learning and the proposed residual NN-based architecture are described in Section 6.4. The experimental setup and the achieved results are discussed in Section

[1] A detailed description of the different ML approaches applied for SIC in FD transceivers can be found in chapter 5.

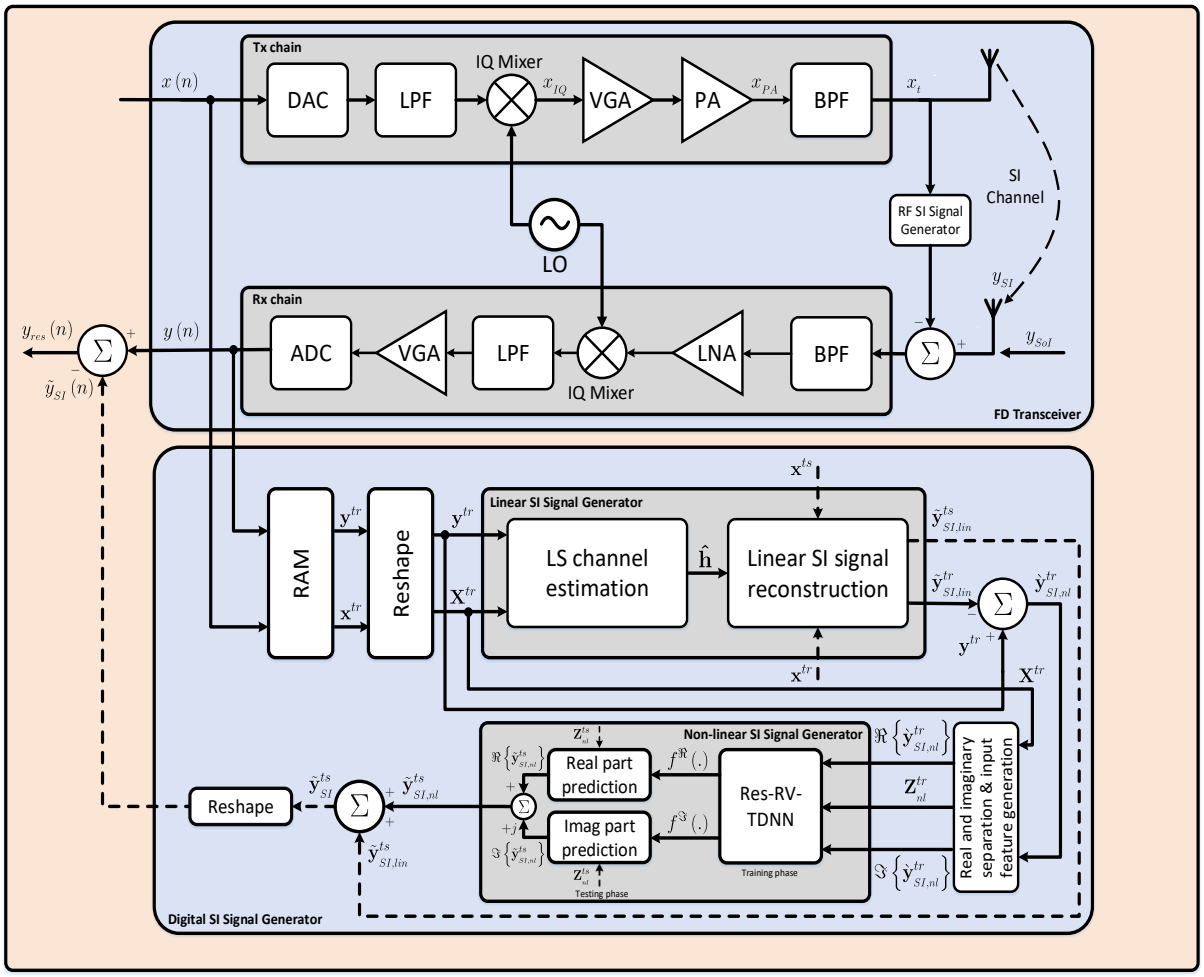


Fig. 6.1: A system model integrating Res-NN with FD communications [18].

6.5. Finally, Section 6.6 concludes the chapter.

6.3 System Model

The system model, integrating Res-NN with FD communications, is shown in Fig. 6.1. The digital baseband signal after the ADC, i.e., after applying the propagation and analog domain cancellation (if any) and upon exposure to different transceiver's impairments, e.g., power amplifier (PA) and I/Q mixer's non-linearities, can be written as [18]

$$y\left(n\right)=y_{_{SI}}\left(n\right)+y_{_{SoI}}\left(n\right)+w\left(n\right), \tag{6.1}$$

with $w\left(n\right)\sim\mathcal{CN}\left(0,\sigma^{2}\right)$ as a complex-valued Gaussian distributed thermal noise with zero mean and variance σ^{2}, $y_{_{SoI}}\left(n\right)$ as the SoI, while $y_{_{SI}}\left(n\right)$ as the SI signal, which can be written as [8]-[13], [18]

$$y_{_{SI}}(n)=\sum_{\substack{p=1,\\ p\,odd}}^{P}\sum_{q=0}^{p}\sum_{m=0}^{M_{i}-1}h_{p,q}\left(m\right)\,x\left(n-m\right)^{q}x^{*}\left(n-m\right)^{p-q}, \tag{6.2}$$

where P denotes the order of non-linearity and $h_{p,q}\left(m\right)$ represents an impulse response of a channel comprising the effect of all transceiver non-idealities [18]. Finally, M_{i} represents the memory effect incorporated into the input samples $x\left(n\right)$ by the SI channel, PA, etc.

For the sake of simplicity, I assume that there is no SoI from any remote TRPs; thus, the aim of the digital canceler is to find an accurate estimate for the SI signal, which I denote by $\widetilde{y}_{_{SI}}(n)$. This can be done by first estimating the linear SI component using the least-squares channel estimation and then estimating the non-linear component using the Res-NN model, as illustrated in Fig. 6.1. The achieved SIC is then quantified over a window of N testing samples as [8]-[13], [18]

$$\mathcal{C}_{dB}=10\log_{10}\left(\frac{\sum_{n=1}^{N}\left|y(n)\right|^{2}}{\sum_{n=1}^{N}\left|y_{_{res}}(n)\right|^{2}}\right), \tag{6.3}$$

with $y_{_{res}}(n)$ as the residual SI signal after applying digital SIC using the linear and non-linear cancelers, as shown in Fig. 6.1.[2]

[2]A detailed description of the FD system model shown in Fig. 6.1 can be found in chapter 5.

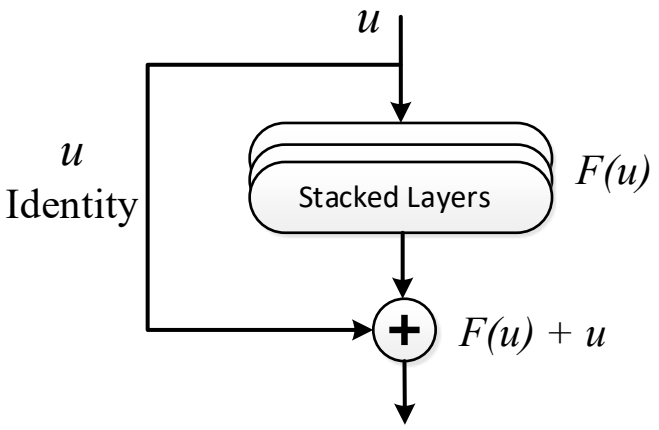


Fig. 6.2: Building structure of Res-NN.

6.4 Proposed Residual NN Architecture

Residual learning of deep NNs [19] was initially proposed to mitigate the performance degradation problem, i.e., increasing the depth of NN leads to an increase in the training errors, which in turn decreases the accuracy of the NN models [20]. Res-NN differs from its NN counterparts by introducing a shortcut connection between the input and output layers of the network, as depicted in Fig. 6.2. This shortcut represents an identity mapping and is added to the output of the stacked/hidden layers. In such manner, the Res-NN drives the non-linear layers to fit a residual mapping $F(u) := H(u) - u$, given a desired underlying mapping $H(u)$, i.e., the original mapping is adjusted to $F(u) + u$.

The adjusted mapping, i.e., $F(u) + u$, is flexible in construction and can be easily realized by different types of feed-forward NNs, and the entire network can eventually be trained by the traditional backpropagation algorithm [19], [20]. In addition, the extra identity shortcut requires only an element-wise addition and, thus, has a negligible effect on the required memory storage and computational complexity compared to the traditional NNs [20], [21]. The aforementioned illustrates the easy implementation and optimization of Res-NN.

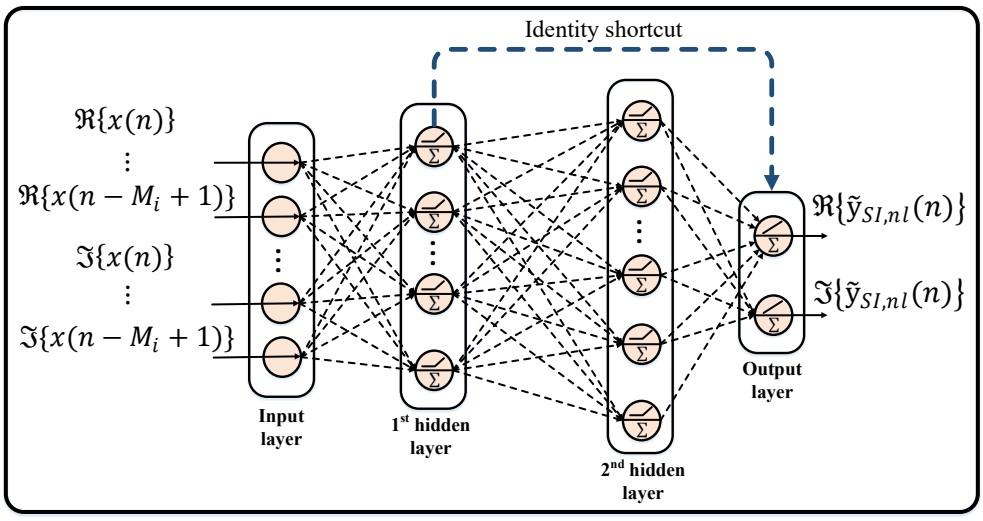


Fig. 6.3: Network architecture of the proposed Res-RV-TDNN.

Inspired by residual learning, I develop a customized Res-NN referred to as residual real-valued time-delay NN (Res-RV-TDNN) to model the FD non-linear SI by exploiting shortcut connections from the first hidden layer to the output layer and thus enhance the learning capabilities of the existing NN-based SI cancelers.

The network design of the proposed Res-RV-TDNN is shown in Fig. 6.3. Herein, I have introduced some modifications to the Res-NN [19]-[21] and the classic RV-TDNN [8] to enhance their learning capabilities for the SIC problem. I construct the combined Res-NN and RV-TDNN with two hidden layers, as follows. The input to the model is a feature map that considers the separated real and imaginary parts of the current and previous samples of the input signal $x\,(n)$. To take the memory effect of the SI into account, the identity connection is taken from the first hidden layer and not from the input as in the original method, i.e., Res-NN. In more detail, I employed an identity shortcut connection to forward a transformed version of the weighted sum of the model inputs (including all the current and past samples) from one neuron of the first hidden layer to the output layer in order to consider the temporal behavior of the SI signal. Moreover, I used non-linear activation functions in the first and second

TABLE 6.1: Training datasets specifications.

Parameter	**First dataset [8]**	**Second dataset [9]**
Modulation	QPSK-OFDM	QPSK-OFDM
FFT size	1024	2048
Pass-band bandwidth	10 MHz	20 MHz
Sampling frequency	20 MHz	80 MHz
Average transmit power	10 dBm	32 dBm
Passive analog SIC	53 dB	15 dB
Active analog SIC	N/A	50 dB
Total analog SIC	53 dB	65 dB
Transmit/receive antennas	1T1R	1T1R
Dataset size	5000 samples	5000 samples
Training/test splits	0.9/0.1	0.9/0.1

hidden layers (i.e., stacked non-linear layers) to fit the adjusted residual mapping, involving the SI non-linearities.

Appending this customized connection to the RV-TDNN model quite enhances it from both performance and complexity perspectives, as will be shown in the following section.

6.5 Results and Discussion

In this section, I compare the performance of the proposed Res-RV-TDNN with that of the best cancelers introduced in [18], i.e., polynomial and RV-TDNN-based cancelers. The comparison between the different cancelers in this study is conducted based on two different datasets that were captured using real-time FD testbeds [8], [9]. The first dataset's average transmit power, bandwidth, and sampling frequency are set to 10 dBm, 10 MHz, and 20 MHz, respectively [8], while those of the second are set to 32 dBm, 20 MHz, and 80 MHz, respectively [9]. Using these setups, 5000 data samples from each dataset are utilized for training and testing. Among the 5000 data samples, the first 90% of the samples are reserved for training, while the last 10% are exploited

TABLE 6.2: Optimal hyperparameters of the proposed Res-RV-TDNN compared to those of the polynomial and RV-TDNN-based cancelers [18].

Dataset	**Canceler**	M_i	P	Nh_1	Nh_2	LR	BS	Act.	Opt.
Fisrt	Polynomial	7	11	-	-	-	-	-	-
	RV-TDNN	9	-	57	-	0.005	22	Relu	Adam
	Res-RV-TDNN	13	-	10	46	0.0025	22	Relu	Adam
Second	Polynomial	9	6	-	-	-	-	-	-
	RV-TDNN	6	-	100	-	0.0025	22	Relu	Adam
	Res-RV-TDNN	6	-	10	48	0.0025	22	Relu	Adam

for testing. The specifications of the two training datasets employed in this study are detailed in Table 6.1.

Based on the above-mentioned datasets, I have optimized the hyperparameters of the proposed Res-RV-TDNN to achieve better or comparable performance to that of the polynomial and RV-TDNN-based cancelers introduced in [18]. In more detail, I have optimized the memory length M_i, the number of neurons in the first and second hidden layers, Nh_1 and Nh_2, the activation function, the learning rate (LR), the batch size (BS), and the training optimizer. The optimal hyperparameters of the proposed Res-RV-TDNN, along with those of the polynomial and RV-TDNN, are summarized in Table 6.2.

Based on the obtained optimal hyperparameters, I have compared, in Fig. 6.4, the SIC performance of the proposed Res-RV-TDNN versus those of the polynomial and RV-TDNN over the first and second datasets, employing 10 dBm and 32 dBm, respectively. As can be seen from the figure, the proposed Res-RV-TDNN achieves a better SIC performance than the RV-TDNN and a comparable SIC to the polynomial in the first dataset. However, it attains a better SIC performance than the polynomial-based canceler and a comparable SIC to the RV-TDNN-based canceler in the second dataset.

The PSDs of the residual SI signal after applying the polynomial-, RV-TDNN-,

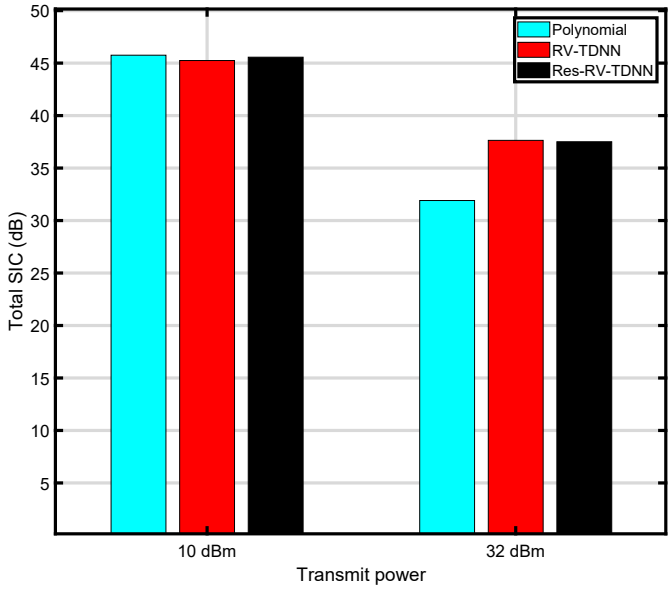


Fig. 6.4: SIC of the proposed Res-RV-TDNN compared to the polynomial and RV-TDNN-based cancelers.

and Res-RV-TDNN-based SIC approaches for the first and second datasets are shown in Figs. 6.5 and 6.6, respectively. From Fig. 6.5, it is observed that the polynomial-, RV-TDNN-, and Res-RV-TDNN-based cancelers provide a comparable gap to Rx noise floor values of 2.1, 2.6, and 2.3 dB, respectively. Further, Fig. 6.6 demonstrates that the polynomial-based canceler provides a large gap to Rx noise floor, i.e., 9.7 dB, compared to the RV-TDNN- and Res-RV-TDNN-based cancelers that only have a gap to Rx noise floor values of 4 and 4.1 dB, respectively. This verifies the prominence of applying data-driven ML techniques to model the SI when high transmit power is employed in FD systems.

The previous results also substantiate the feasibility of adding residual connections to the NN, as it guarantees a good SIC performance, regardless of the transmit power level, i.e., the Res-RV-TDNN works well in the case of low transmit power levels (first dataset) and also for the case of the higher transmit power levels (second dataset).

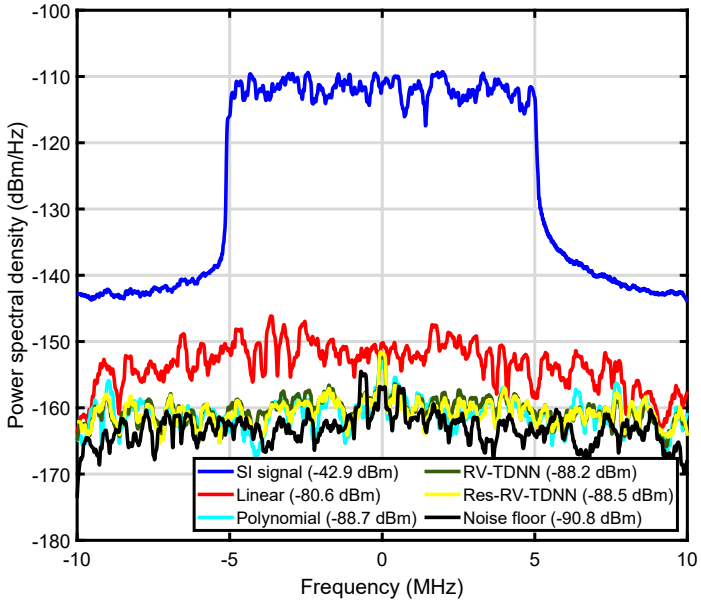


Fig. 6.5: PSD of the proposed Res-RV-TDNN compared to the polynomial and RV-TDNN-based cancelers over the first dataset.

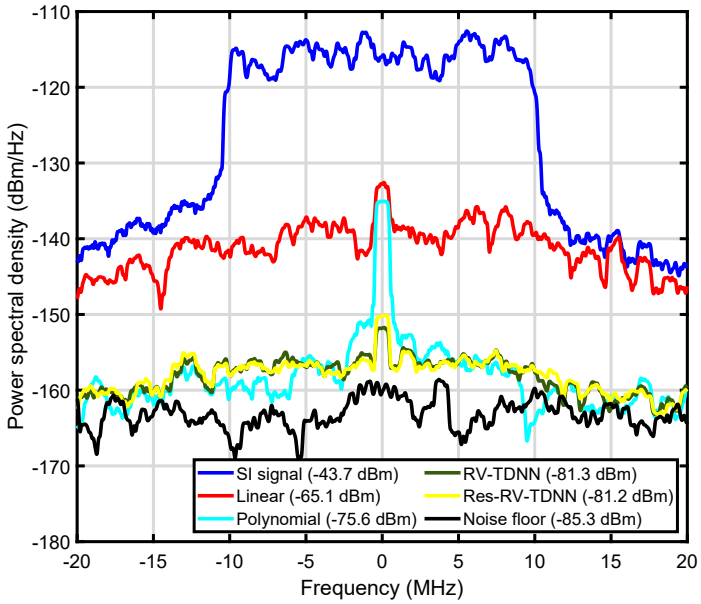


Fig. 6.6: PSD of the proposed Res-RV-TDNN compared to the polynomial and RV-TDNN-based cancelers over the second dataset.

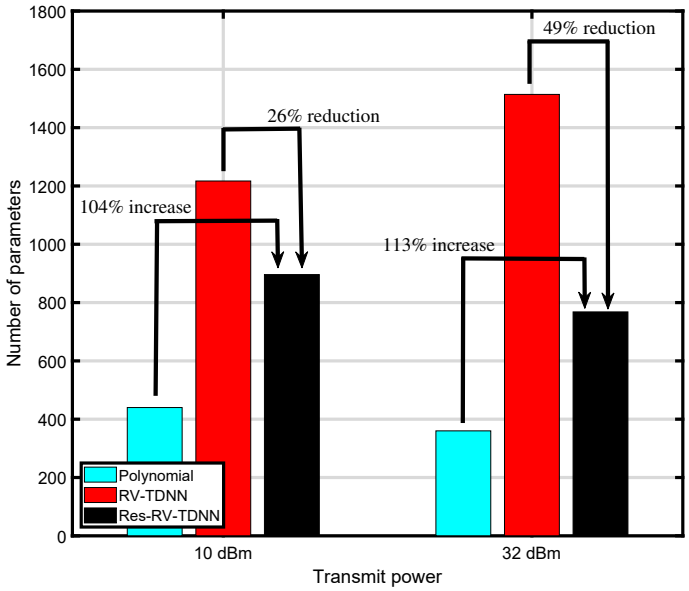


Fig. 6.7: Number of parameters of the proposed Res-RV-TDNN compared to the polynomial and RV-TDNN-based cancelers.

Upon evaluating the SIC and PSD performances, I have compared the required memory storage—in terms of the number of parameters—of the proposed Res-RV-TDNN compared to that of the polynomial and RV-TDNN-based cancelers over the first and second datasets, in Fig. 6.7. As can be seen from the figure, the proposed Res-RV-TDNN provides up to 26% and 49% reduction in the number of required parameters compared to the RV-TDNN in the first and second datasets, respectively. However, it demands approximately double the parameters required by the polynomial-based canceler in both datasets.

In Fig. 6.8, I have evaluated the computational complexity of the proposed Res-RV-TDNN—in terms of the number of floating points operations (FLOPs) in the inference phase—and compared it with that of the polynomial and RV-TDNN-based cancelers over the first and second datasets. As can be seen from the figure, the proposed Res-RV-TDNN significantly reduces the number of FLOPs compared to the

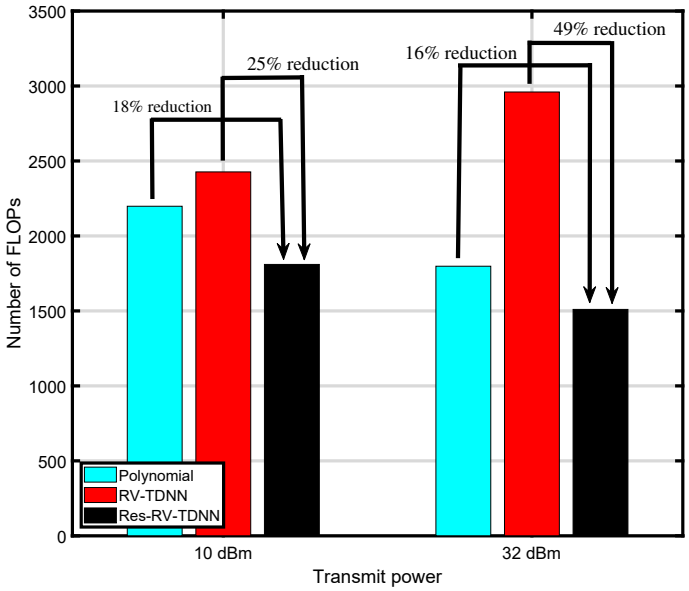


Fig. 6.8: Number of FLOPs of the proposed Res-RV-TDNN compared to the polynomial and RV-TDNN-based cancelers.

polynomial and RV-TDNN-based cancelers for the two datasets. In more detail, it provides a 25% and 49% reduction in the computational complexity over the RV-TDNN in the first and second datasets, respectively. In addition, it exhibits about a 18% and 16% reduction in the complexity compared to the polynomial canceler for both datasets, respectively. The previous results substantiate the superiority of the proposed Res-RV-TDNN in achieving a high SIC with low computational requirements.

Upon evaluating the SIC, memory storage, and computational complexity of various SI cancelers, I have evaluated the efficiency η of each canceler for different test cases/system requirements in Table 6.3, as in [18].[3] From the results in Table 6.3, it is observed that the proposed Res-RV-TDNN surpasses the classical RV-TDNN for both datasets. In addition, it can be concluded that the polynomial canceler is the

[3]In this chapter, the canceler efficiency η of each SI canceler is evaluated in terms of its SIC performance, memory storage, and computational complexity. A detailed description of calculating η can be found in [18].

TABLE 6.3: Efficiency η of RV-TDNN and Res-RV-TDNN-based SI cancelers compared to the polynomial canceler for the first and second datasets.

Test case	η					
	First dataset			**Second dataset**		
	Polynomial	RV-TDNN	Res-RV-TDNN	Polynomial	RV-TDNN	Res-RV-TDNN
SIC is the only system criterion.	$\mathbf{1^{st}}$	3^{rd}	2^{nd}	3^{rd}	$\mathbf{1^{st}}$	2^{nd}
SIC and memory are the only system criteria.	$\mathbf{1^{st}}$	3^{rd}	2^{nd}	3^{rd}	2^{nd}	$\mathbf{1^{st}}$
SIC and complexity are the only system criteria.	2^{nd}	3^{rd}	$\mathbf{1^{st}}$	3^{rd}	2^{nd}	$\mathbf{1^{st}}$
SIC, memory, and complexity are the only system criteria.	$\mathbf{1^{st}}$	3^{rd}	2^{nd}	3^{rd}	2^{nd}	$\mathbf{1^{st}}$

The cancelers that achieve the highest efficiency are ranked by the 1^{st}, 2^{nd}, and 3^{rd}, respectively.

efficient one when using the first dataset, while the proposed Res-RV-TDNN is the efficient canceler when using the second dataset; note that the second dataset uses higher transmission power, bandwidth, and sampling frequency hence considered a more realistic representation for the TRPs, employing FD communications.

6.6 Conclusion

In this chapter, I proposed a residual NN-based SI canceler, referred to as Res-RV-TDNN, to effectively learn the FD SI problem. The Res-RV-TDNN combined the concepts of residual learning and time-delay NNs to model the memory effect and different non-linearities affecting the SI in FD transceivers. The proposed method is examined by using two training datasets and various performance metrics such as digital SIC, memory requirements, and computational complexity. The achieved results demonstrate that the proposed method gives a similar non-linear SIC to the existing benchmarks in the literature and provides significant reductions in computational resources. Further, the proposed method is shown to be more efficient for realistic FD operating scenarios, employing high transmit power levels, bandwidth, and sampling frequency. Exploiting the concept of residual learning to further enhance the performance of grid-based, hybrid-layers, and output feedback NN structures can be considered in the future.

References

[1] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

[2] F. Abbas, X. Yuan, M. S. Bute, and P. Fan, "Performance analysis using full duplex discovery mechanism in 5G-V2X communication networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 11453–11464, Aug. 2022.

[3] Z. Wang, W. Shi, W. Liu, Y. Zhao, and K. Kang, "Performance analysis of two-way full-duplex relay mixed RF/FSO system with self-interference," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 209–213, Jan. 2021.

[4] L. Shen, B. Henson, Y. Zakharov, and P. Mitchell, "Digital self-interference cancellation for underwater acoustic systems," *IEEE Trans. Circuits Syst. II: Exp. Briefs*, vol. 67, no. 1, pp. 192–196, Jan. 2020.

[5] 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.

[6] 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.

[7] D. Korpi, L. Anttila, and M. Valkama, "Nonlinear self-interference cancellation in MIMO full-duplex transceivers under crosstalk," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, pp. 1–15, Dec. 2017.

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

[9] 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.

[10] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Advanced machine learning techniques for self-interference cancellation in full-duplex radios," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 1149–1153.

[11] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Low complexity neural network structures for self-interference cancellation in full-duplex radio," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 181–185, Jan. 2021.

[12] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Full-duplex self-interference cancellation using dual-neurons neural networks," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 557–561, Mar. 2022.

[13] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Hybrid-layers neural network architectures for modeling the self-interference in full-duplex systems," *IEEE Trans. Veh Technol.*, vol. 71, no. 6, pp. 6291–6307, Jun. 2022.

[14] M. Erdem, H. Ozkan, and O. Gurbuz, "Nonlinear digital self-interference cancellation with SVR for full duplex communication," in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC)*, Jun. 2020, pp. 1–6.

[15] M. Yilan, O. Gurbuz, and H. Ozkan, "Integrated linear and nonlinear digital cancellation for full duplex communication," *IEEE Wireless Communications*, vol. 28, no. 1, pp. 20–27, Feb. 2021.

[16] F. Jochems and A. Balatsoukas-Stimming, "Non-linear self-interference cancellation via tensor completion," in *Proc. 54th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2020, pp. 905–909.

[17] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Identification of nonlinear RF systems using backpropagation," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2020, pp. 1–6.

[18] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Machine learning-based self-interference cancellation for full-duplex radio: Approaches, open challenges, and future research directions," *IEEE OJVT*, early access, pp. 1–25, Nov. 2023.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Jun. 2016, pp. 770–778.

[20] H. Zhu, Z. Cao, Y. Zhao, and D. Li, "Learning to denoise and decode: A novel residual neural network decoder for polar codes," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 8725–8738, Aug. 2020.

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

Chapter 7

Extreme Learning Machine-Assisted Full-Duplex Self-Interference Cancellation

7.1 Abstract

This chapter introduces an extreme learning machine (ELM)-assisted self-interference cancellation (SIC) approach to suppress the self-interference signal in full-duplex transceivers. The proposed ELM-assisted SIC approach exploits a single hidden layer neural network (NN) topology without performing iterative backpropagation tuning, yielding reduced training time and better generalization than the conventional NN-assisted SIC approaches. The over-the-air measurements corroborate that the proposed ELM-assisted approach could achieve a higher SIC performance with a lower training time than the literature benchmarks, albeit at higher memory and computational requirements.

7.2 Introduction

Full-duplex (FD), enabling remote parties to transfer information simultaneously in both directions and in the same bandwidth, has been envisioned as an important technology for the next-generation cellular networks [1]. This is due to the ability to leverage both time and frequency resources and theoretically double the spectral efficiency. Enabling the FD communications is, however, highly challenging due to the self-interference (SI) [2]. The power of the SI signal is significantly higher when compared with the signal of interest (SoI) from a remote node as a result of the proximity of the Tx to its co-located receiver (Rx). The SI signal is thus swamping the SoI and degrading the FD system's overall performance.

In the past, traditional self-interference cancellation (SIC) approaches, spanning the propagation, analog, and/or digital techniques, have been explored to diminish the SI. Such approaches are shown to be effective for SIC in FD transceivers (TRXs); however, they could impose additional cost, hardware, memory, and/or computational complexity requirements [3].

In recent years, applying machine learning (ML)-assisted SIC approaches for suppressing the SI in FD TRXs has gained popularity to address the extra requirements imposed by the traditional methods [4]-[11]. Diverse neural networks (NNs) [4]-[8] and other ML approaches, such as support vector regressors [9], tensor completion [10], and more [11], have been extensively explored to reduce the FD SI signal's power with lower cost, hardware, memory, and/or computational requirements. Applying ML-assisted SIC approaches for suppressing the FD SI, however, can come at the expense of excessive training time, especially in the case of NNs, tensor completion, etc., making them unfavorable for real-time dynamic scenarios where the training time is a crucial factor [11]. To cope with this problem, this chapter proposes an extreme learning machine (ELM)-assisted SIC approach for suppressing the SI in FD

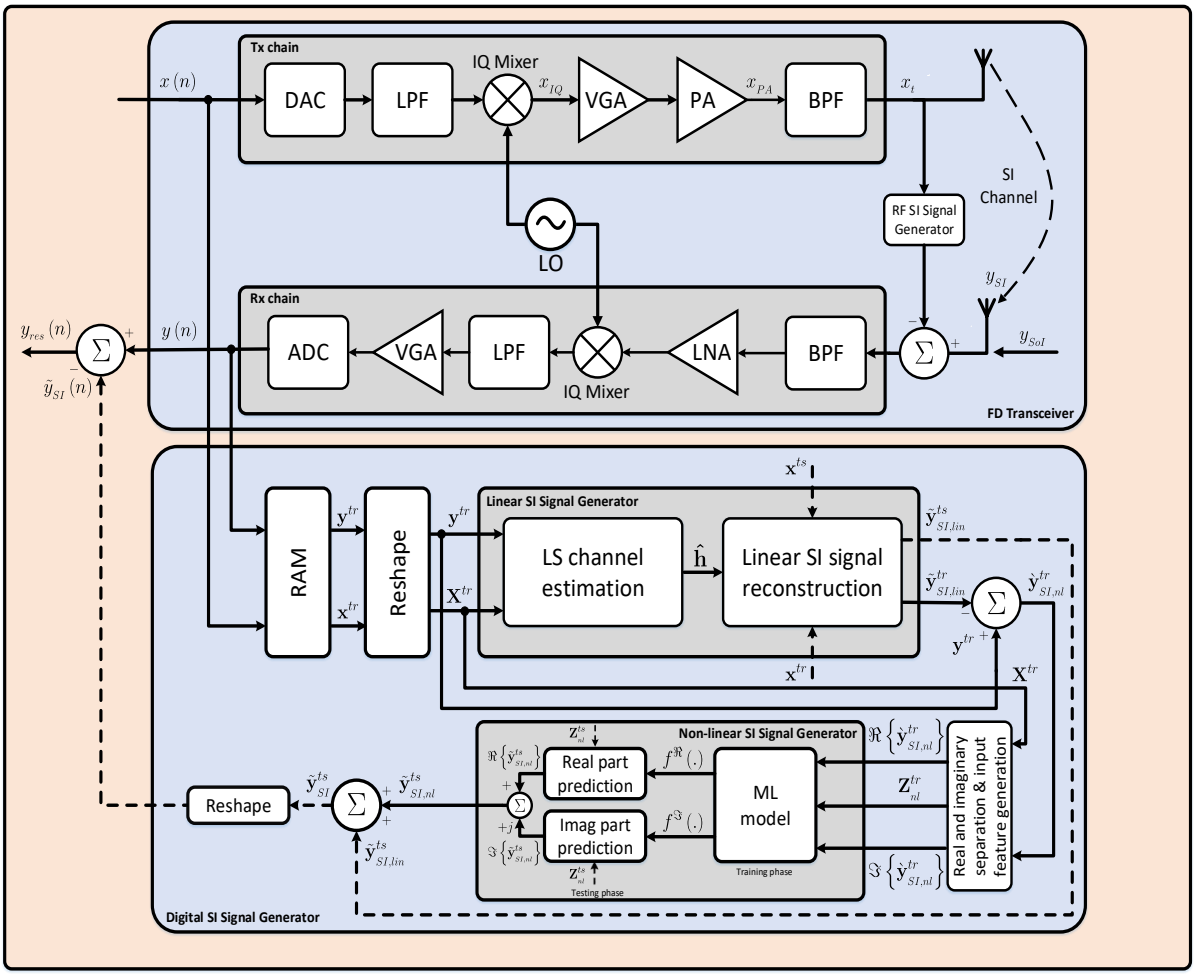


Fig. 7.1: ML-assisted FD system model [11].

TRXs with higher SIC performance and lower training overhead compared to the existing benchmarks. In addition, this chapter conducts comprehensive evaluations of the proposed SIC approach and current benchmarks using captured data from real-time FD experiments, which are found to be significant for performance verification of real-world communication systems [12].

The original contributions brought by this chapter are listed as follows:

1. I propose an ELM-assisted SIC approach for suppressing the SI in FD TRXs, for the first time in literature.

2. Using the over-the-air measurements, I evaluate the SIC performance of the

ELM-assisted SIC approach and compare it to the existing benchmarks, employing polynomial and NN modeling.

3. I also investigate the training overhead, memory resources, and computational requirements of the ELM-assisted SIC approach and compare them with the benchmarks.

This chapter is structured as follows: A general ML-assisted FD system model is outlined in Section 7.3. The proposed ELM-assisted SIC approach is presented in Section 7.4. Numerical results are discussed in Section 7.5, and conclusions are drawn in Section 7.6.

7.3 ML-Assisted FD System Model

A generic ML-assisted FD system model is shown in Fig. 7.1. The input samples, $x(n)$, with n denoting the sample's index, are exposed to various impairments of the TRX's blocks, such as the power amplifier (PA) and analog-digital/digital-to-analog domain converters (ADC and DAC). The baseband samples after digitization are given by [4]-[8], [11]

$$y(n) = y_{SI}(n) + y_{SoI}(n) + \psi(n), \tag{7.1}$$

where $\psi(n) \sim \mathcal{CN}(0, \sigma^2)$ indicates the thermal noise samples, having a complex-valued zero-mean normal distribution with variance σ^2. $y_{SoI}(n)$ represents the SoI, whereas $y_{SI}(n)$ denotes the SI signal, which can be expressed as [4]-[8], [11]

$$y_{SI}(n) = \sum_{\substack{p=1,\\ p\,odd}}^{P} \sum_{q=0}^{p} \sum_{m=0}^{M_i-1} h_{p,q}(m)\, x(n-m)^q\, x^*(n-m)^{p-q}, \tag{7.2}$$

with $h_{p,q}(m)$ as the channel impulse response modeling the impact of the TRX's impairments, P as the PA's non-linearity order, and M_i as the memory length associated with the input samples, respectively [11].

I assume that no SoI is received from any other remote nodes, for the ease of exposition [4]-[8], [11]. The task of the digital canceler is thus to generate a proper estimate of the SI signal $y_{SI}(n)$, represented by $\widetilde{y}_{SI}(n)$. This can be achieved through two main tasks, as depicted in Fig. 7.1. Firstly, the common least-squares channel estimation technique is used to generate the linear component of the SI. Then, the non-linear component is generated using any ML model, e.g., NN, ELM. The total attainable SIC is then evaluated over a number of testing samples N as [4]-[8], [11]

$$\mathcal{C}_{dB} = 10\log_{10}\left(\left(\sum_{n=1}^{N}|y(n)|^2\right)\left(\sum_{n=1}^{N}|y_{res}(n)|^2\right)^{-1}\right), \tag{7.3}$$

where $y_{res}(n)$ represents the residual SI after the application of the digital SIC techniques, as illustrated in Fig. 7.1.[1]

7.4 Proposed ELM-Assisted SIC Approach

ELM is a fast and accurate learning algorithm that relies on a single hidden layer feed-forward NN topology [13], [14]. In ELM, the hidden nodes' parameters, i.e., weights and biases, are not traditionally tuned using the iterative backpropagation algorithm as in the conventional NNs; however, they are randomly assigned from a certain distribution and then utilized to analytically compute the output layer's parameters, i.e., output layer's weights [13], [14]. Avoiding backpropagation tuning enables the ELM to achieve a better generalization with a lower training time than

[1]The ML-assisted FD system model, depicted in Fig. 7.1, is detailed in [11].

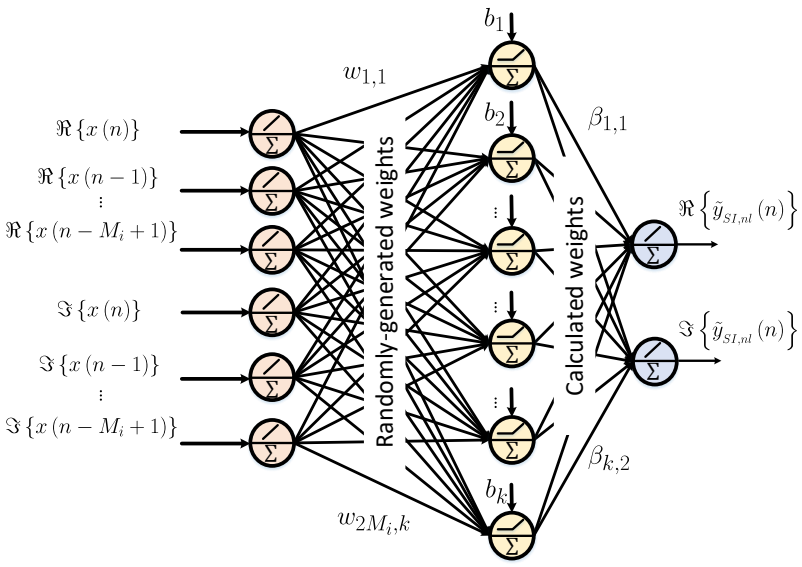


Fig. 7.2: Proposed ELM-assisted SIC approach.

$$\mathbf{Z}_{nl}^{tr} = \begin{bmatrix} \Re\{x(n)\} & \dots & \Re\{x(n-M_i+1)\} & \Im\{x(n)\} & \dots & \Im\{x(n-M_i+1)\} \\ \Re\{x(n+1)\} & \dots & \Re\{x(n-M_i+2)\} & \Im\{x(n+1)\} & \dots & \Im\{x(n-M_i+2)\} \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \Re\{x(n+N_{tr}-M_i-1)\} & \dots & \Re\{x(n+N_{tr}-2M_i)\} & \Im\{x(n+N_{tr}-M_i-1)\} & \dots & \Im\{x(n+N_{tr}-2M_i)\} \end{bmatrix}, \tag{7.4}$$

the traditional NNs [13], [14]. Inspired by these potentials, this chapter proposes an ELM-assisted SIC approach for modeling the SI in FD TRXs. The architecture of the designed ELM is illustrated in Fig. 7.2. According to the figure, the proposed ELM exploits an NN topology consisting of three layers, namely the input, hidden, and output layers. The feature map $\mathbf{Z}_{nl}^{tr} \in \mathbb{R}^{(N_{tr}-M_i)\times 2M_i}$, containing the in-phase, $(\Re(.))$, and quadrature, $(\Im(.))$, components of both the current and past (memory effect) input samples, is utilized as an input for training, as expressed in (7.4), with $\mathbb{R}$ as the set of real-valued numbers and N_{tr} as the number of training samples. At the hidden layer, the hidden nodes' parameters are randomly generated from a standard normal distribution. The output matrix, $\mathbf{H} \in \mathbb{R}^{(N_{tr}-M_i)\times k}$, corresponding to the hidden layer, is then computed as [13], [14]

$$\mathbf{H} = \begin{bmatrix} f\left(\mathbf{z}_{nl,1}^{tr}\mathbf{w}_1 + b_1\right) & \cdots & f\left(\mathbf{z}_{nl,1}^{tr}\mathbf{w}_k + b_k\right) \\ f\left(\mathbf{z}_{nl,2}^{tr}\mathbf{w}_1 + b_1\right) & \cdots & f\left(\mathbf{z}_{nl,2}^{tr}\mathbf{w}_k + b_k\right) \\ \ddots & \ddots & \ddots \\ f\left(\mathbf{z}_{nl,N_{tr}-M_i}^{tr}\mathbf{w}_1 + b_1\right) & \cdots & f\left(\mathbf{z}_{nl,N_{tr}-M_i}^{tr}\mathbf{w}_k + b_k\right) \end{bmatrix}, \tag{7.5}$$

with $f(\centerdot)$ as the activation function operator and k as the number of ELM hidden neurons. $\mathbf{z}_{nl,1}^{tr},\cdots,\mathbf{z}_{nl,N_{tr}-M_i}^{tr} \in \mathbb{R}^{1\times 2M_i}$ represent the row vectors of the predefined $\mathbf{Z}_{nl}^{tr}$, while $\mathbf{w}_1,\cdots,\mathbf{w}_k \in \mathbb{R}^{2M_i\times 1}$ denote the column vectors of the randomly-generated weight matrix $\mathbf{W} \in \mathbb{R}^{2M_i\times k}$. Finally, $b_1,\cdots,b_k \in \mathbb{R}^{1\times 1}$ indicate the bias elements drawn from the randomly-generated bias vector $\mathbf{b} \in \mathbb{R}^{k\times 1}$. The randomly-generated weight matrix $\mathbf{W}$ and bias vector $\mathbf{b}$ can be respectively expressed as

$$\mathbf{W} = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,k} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,k} \\ \ddots & \cdots & \cdots & \ddots \\ w_{2M_i,1} & w_{2M_i,2} & \cdots & w_{2M_i,k} \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \ddots \\ b_k \end{bmatrix}. \tag{7.6}$$

The output layer's parameters, i.e., output layer's weights, $\boldsymbol{\beta} \in \mathbb{R}^{k\times 2}$, are then learned, being computed as [13], [14]

$$\boldsymbol{\beta} = \mathbf{H}^{\dagger}\grave{\mathbf{y}}_{SI,nl}^{tr}, \tag{7.7}$$

where $(\centerdot)^{\dagger}$ indicates the Moore–Penrose inverse operation and $\grave{\mathbf{y}}_{SI,nl}^{tr} \in \mathbb{R}^{(N_{tr}-M_i)\times 2}$ denotes the output vector containing the training labels for the ELM model, which are constructed using the residual SI after performing the linear cancellation, as shown in Fig. 7.1. Upon finishing the training process, the trained ELM model is fed with the testing matrix $\mathbf{Z}_{nl}^{ts}$ to generate the estimated SI signal's samples $\widetilde{\mathbf{y}}_{SI,nl}^{ts}$, as depicted in Fig. 7.1, with N_{ts} representing the number of testing samples.

7.5 Numerical Results

Here, I compare the performance of the ELM-assisted SIC approach with the existing benchmarks, namely polynomial and RV-TDNN—a time-delay NN based on a real-valued framework [4], [11]. The measurement setup, hyperparameter tuning, and performance metrics used to assess the aforementioned approaches are detailed in the following subsections.

7.5.1 Measurement Setup

The experimental setup employed for capturing the over-the-air measurements utilized to train the proposed ELM-assisted SIC approach is described in this subsection. The measured data was collected using a real-time FD testbed [4], [10]; this relies on a National Instruments PXI platform and a National Instruments NI-5791 RF card for data transmission and reception. An orthogonal frequency division multiplexing signal, modulated by a QPSK technique, was generated over 20 MHz bandwidth and sampled at an 80 MHz sampling rate. A 2.45 GHz carrier frequency and a 32 dBm average transmit power were set. At the Rx side, 15 dB and 50 dB passive and active analog cancellations were respectively used, resulting in 65 dB total analog cancellation. Using the above setup, 20,480 data samples were stored in a dataset to train the ELM-assisted SIC approach [4], [10]. Among the 20,480 samples, 90% were reserved for training, whereas 10% were customized for testing. Upon capturing the training dataset, 3.7.5 and 5.1.5 versions of Python and Spyder were respectively used in a Windows environment for post-processing, i.e., generating the cancellation signals after performing the training for the ELM-assisted SIC approach.[2]

[2]The measurement setup described in Section 7.5.1 is detailed in [4], [11].

TABLE 7.1: Optimal hyperparameters of the proposed ELM-, polynomial-, and RV-TDNN-assisted SIC approaches.

SI Canceler	M_i	P	k	LR	BS	Act.	Opt.
Polynomial	13	9	-	-	-	-	-
RV-TDNN	6	-	53	0.005	22	Relu	Adam
ELM	6	-	702	-	-	tanh	-

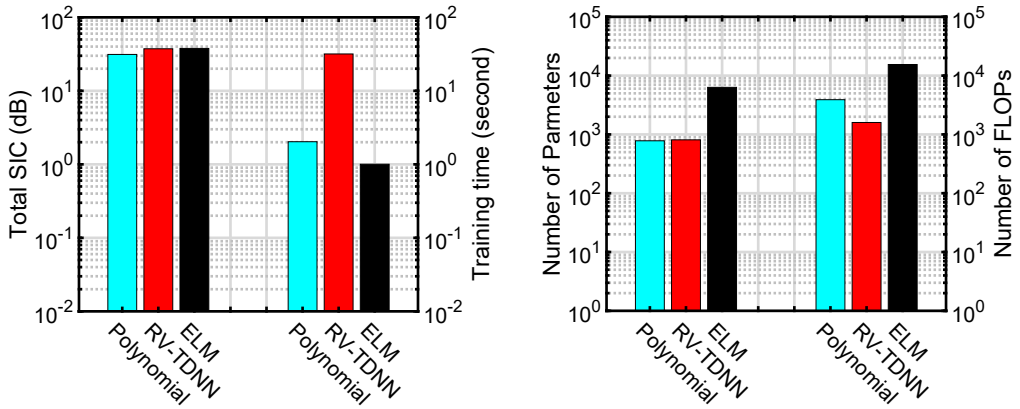
7.5.2 Hyperparameter Tuning

A hyperparameter tuning for the proposed ELM-assisted SIC approach is performed to achieve its maximum SIC. The obtained SIC is then compared with the maximum SIC acquired by the polynomial and the RV-TDNN approaches [4]. Specifically, I have optimized the memory length, number of hidden neurons, and activation function by considering $M_i \in \{1, 2, \ldots, 13\}$, $k \in \{1, 2, \ldots, 2000\}$, and $f \in \{Relu, Sigmoid, tanh\}$, respectively. The optimal hyperparameters for the proposed ELM-, polynomial-, and RV-TDNN-assisted SIC approaches are summarized in Table 7.1.

7.5.3 Performance Metrics

7.5.3.1 SIC Performance

The attainable performance of the proposed ELM-assisted SI canceler is compared to that accomplished by the polynomial- and RV-TDNN, as illustrated in Fig. 7.3a. The ELM-assisted canceler outperforms the two benchmarks by achieving a higher SIC performance. The previous results have also been confirmed in Fig. 7.4 by plotting the power spectral density (PSD) of the residual SI after the application of the aforementioned non-linear cancellation using the above techniques. As can be observed from Fig. 7.4, the ELM-assisted canceler efficiently reduces the power of the SI signal to the Rx noise floor level by providing a gap to the Rx's noise floor of 4.3 dB compared to 10.8 dB and 4.7 dB achieved by the polynomial and RV-TDNN, respectively.



(a) SIC & training time.

(b) Parameters & FLOPs.

Fig. 7.3: Performance analysis of the proposed ELM-assisted SIC approach compared to polynomial and RV-TDNN-assisted SIC approaches.

7.5.3.2 Training Overhead

The training overhead, i.e., training time, of the proposed ELM-assisted SI canceler compared to the polynomial and RV-TDNN-assisted cancelers is also depicted in Fig. 7.3a. As seen from the figure, the proposed ELM distinctly reduces the training time; this is a result of performing the training process without conducting iterative backpropagation tuning. It is worth mentioning that all ML-assisted SIC approaches reported in the literature have a higher training time than the polynomial-assisted SIC approach [11]; however, the proposed ELM is the only ML method that is able to shorten the training overhead compared to the analytical technique, i.e., the polynomial-based approach, due to its simple mechanism compared to others, as mentioned before.

7.5.3.3 Memory and Computational Requirements

The memory and computational requirements of the proposed ELM-assisted SIC canceler are evaluated in the inference stage by counting the required number of parame-

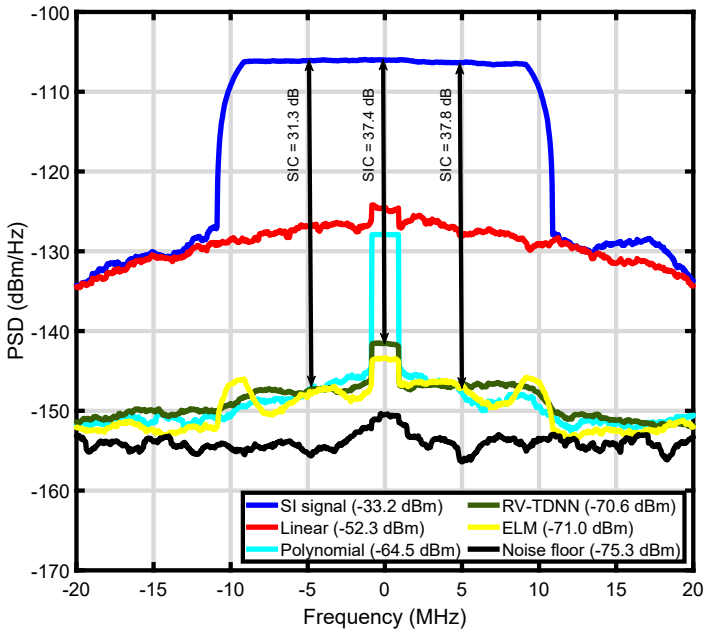


Fig. 7.4: PSD of the proposed ELM-assisted SIC approach compared to polynomial and RV-TDNN-assisted SIC approaches.

ters and floating-point operations (FLOPs), respectively, and compared with those of the polynomial- and RV-TDNN-assisted cancelers in Fig. 7.3b. As can be seen from the figure, the proposed ELM requires a higher number of parameters and FLOPs than the polynomial and RV-TDNN due to the use of a higher number of hidden neurons, as illustrated in Table 7.1.

In summary, the proposed ELM-assisted SIC approach achieves a higher SIC with a lower training overhead than the existing literature benchmarks at the expense of higher memory and computational requirements.

7.6 Conclusion

This chapter proposed an ELM-assisted SIC approach to model the FD SI. The proposed model uses a single hidden layer NN topology, without performing iterative

backpropagation tuning. The performance of the ELM-assisted SIC approach was evaluated and compared with the existing benchmarks using over-the-air measurements obtained from a real-time FD testbed. The over-the-air measurements substantiated that the proposed ELM-assisted SIC approach could achieve a higher SIC with a lower training overhead than the literature benchmarks, albeit with higher memory and computational resources. Solutions to relax such requirements should be sought in the future.

References

[1] L. Bariah *et al.*, "A prospective look: Key enabling technologies, applications and open research topics in 6G networks," *IEEE Access*, vol. 8, pp. 174792–174820, Aug. 2020.

[2] A. M. M. Chandran and M. Zawodniok, "Measurement of internal self-interference of USRP family of devices in full-duplex operations," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (IMTC)*, Jun. 2020, pp. 1-6.

[3] 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.

[4] 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.

[5] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Low complexity neural network structures for self-interference cancellation in full-duplex radio," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 181–185, Jan. 2021.

[6] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Full-duplex self-interference cancellation using dual-neurons neural networks," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 557–561, Mar. 2022.

[7] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Hybrid-layers neural network architectures for modeling the self-interference in full-duplex systems," *IEEE Trans. Veh Technol.*, vol. 71, no. 6, pp. 6291–6307, Jun. 2022.

[8] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Residual neural networks for learning the full-duplex self-interference," in *Proc. 57th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2023, pp. 1–4.

[9] M. Yilan, O. Gurbuz, and H. Ozkan, "Integrated linear and nonlinear digital cancellation for full duplex communication," *IEEE Wireless Communications*, vol. 28, no. 1, pp. 20–27, Feb. 2021.

[10] F. Jochems and A. Balatsoukas-Stimming, "Non-linear self-interference cancellation via tensor completion," in *Proc. 54th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2020, pp. 905–909.

[11] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Machine learning-based self-interference cancellation for full-duplex radio: Approaches, open challenges, and future research directions," *IEEE Open J. Veh. Technol.*, vol. 5, no. 1, pp. 21–47, Nov. 2023.

[12] L. Angrisani, D. Petri, and M. Yeary, "Instrumentation and measurement in communication systems," *IEEE Instrum. Meas. Mag.*, vol. 18, no. 2, pp. 4–10, Apr. 2015.

[13] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.

[14] J. Tang, C. Deng, and G. B. Huang, "Extreme learning machine for multilayer perceptron," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 4, pp. 809–821, Apr. 2016.

Chapter 8

Conclusions and Future Work

In this chapter, I summarize the contributions presented in this thesis and provide a guide for future research directions.

8.1 Conclusions

The focus of this thesis is to apply machine learning (ML)-assisted self-interference cancellation (SIC) approaches to cancel the self-interference (SI) in full-duplex (FD) transceivers. In Chapter 2, two grid-based neural network (NN) structures, referred to as ladder-wise grid structure and moving-window grid structure, have been proposed to model the SI in FD transceivers with lower memory and computational requirements than the literature benchmarks. Further reduction in the computational complexity is provided in Chapter 3, where two hybrid-layers NN structures, referred to as hybrid-convolutional recurrent NN and hybrid-convolutional recurrent dense NN, have been proposed to model the FD SI. The proposed NN structures exhibit lower computational requirements than the grid-based structures and without degradation in the SIC performance. In Chapter 4, an output-feedback NN structure, referred to as the dual neurons-ℓ hidden layers NN, has been designed to model the SI in FD

transceivers with less memory and computational requirements than the grid-based and hybrid-layers NN structures and without any additional deterioration to the cancellation performance. In Chapter 5, support vector regressors (SVRs), variants of support vector machines, have been proposed to cancel the SI in FD transceivers. A case study to assess the performance of SVR-based approaches compared to the classical and other ML-based approaches, using different performance metrics and two different test setups, has also been provided in this chapter. The SVR-based SIC approaches were able to reduce the training time compared to the NN-based approaches, which have, contrarily, shown to be more efficient in terms of SIC, especially when high transmit power is employed. In Chapter 6, the concept of residual learning has been exploited to develop an NN structure, referred to as residual real-valued time-delay NN, to model the FD SI with lower computational requirements than the benchmarks of Chapter 5. In Chapter 7, a fast and accurate learning algorithm, namely extreme learning machine, has been proposed to suppress the SI in FD transceivers with higher SIC performance and lower training overhead than the benchmarks of Chapter 5.

8.2 Potential Directions of Future Investigation

The research presented in this thesis serves as a foundation for prospective investigations, pointing towards several future directions, among which I list the following:

- **Considering the Effect of Signal of Interest (SoI) while Performing the SIC:** The existing ML-based SIC approaches consider the cancellation of the SI signal only, i.e., no signal from any remote FD or half-duplex node is considered. However, in practical situations, i.e., real-time FD systems, the SIC in one FD node has to be done while an SoI from another node is received and demodulated. Initial works in [1], [2] investigated a joint detection of the SI and SoI and proved

that an NN-based SI canceler is beneficial to enhance the signal demodulation. Despite the potential of the works in [1], [2], there are still more issues remaining to be addressed, and the point of detecting the SoI while performing the SIC is open to improvements in both performance and complexity perspectives. For instance, one issue is that all ML-based approaches introduced in this thesis are trained and verified using time-domain samples, i.e., they are completely working in the time domain. However, if the SoI employs any of the frequency-domain modulation formats, e.g., orthogonal frequency division multiplexing, performing the SIC could be done in the frequency domain [1]. Thus, adapting the ML-based SIC approaches to work with frequency- rather than time-domain samples can be a direction for future investigation.

- **Tackling the Time-Varying SI Channels:** The existing ML-based SIC approaches use offline-trained ML algorithms to estimate the SI signal over a static SI channel. However, in practical situations, the movements of user equipment and/or environmental changes can vary the SI channel over time, and the ML algorithms need to be retrained in order to adapt to the time-varying SI channel. Nevertheless, some ML algorithms require a higher training time, i.e., they are not fast enough to be retrained during the FD transmission, which can lead to significant performance degradation. Initial works in [3], [4] investigate the effect of canceling the SI signal under time-varying SI channels. However, these are incipient works, and the point is open to improvements in both performance and complexity perspectives. For instance, applying reinforcement and online learning to iteratively tackle the time-varying SI channel can be a future direction of investigation. Scaling the performance and/or complexity as a result of applying reinforcement and online learning can also be considered in the future.

- **Applying ML Approaches for SIC in FD MIMO Systems:** The ML-based SIC approaches introduced in this thesis are trained and verified using a single-input single-output (SISO) FD testbed. However, in recent communication standards, the multiple-input multiple-output (MIMO) technology has become a basic transmit/receive scheme. Hence, extending the ML-based SIC techniques to MIMO rather than SISO FD transceivers is imperative. Typically, the complexity of the SIC approaches exponentially increases under MIMO operation where M transmit antennas interfere with N receive antennas. A straightforward approach—to process several SI signals in the digital domain—is to perform the SIC using separate SI cancelers, which consider the interfering signals from all transmit antennas; however, this can result in excessive complexity. To address this issue, alternative approaches can be designed. For instance, exploiting the spatial correlation between the MIMO channels to develop a common SI canceler, i.e., not separate cancelers, in order to reduce the impractical computational complexity of the traditional MIMO SIC-based approaches, can be a direction for future investigation [5].

- **Training Complexity of ML-based SIC Approaches:** The computational complexity of the existing ML-based SIC approaches is typically evaluated and compared in terms of the number of floating-point operations (FLOPs) required in the inference stage, i.e., upon performing and finalizing the training process. However, estimating the training complexity (in terms of FLOPs) is crucial and should be considered, especially for ML algorithms targeted to be integrated with reinforcement and online learning. For instance, calculating the number of FLOPs required for performing the backpropagation in NNs and approximating the unknown function using optimization in support vector regressors should be explored to provide insights about the feasibility of applying ML-based ap-

proaches for SIC in real-time FD transceivers.

The aforementioned directions aim to build upon the foundations laid by this dissertation, providing a guide for future exploration in order to successfully integrate data-driven ML approaches with FD communication systems, applying digital SIC.

References

[1] A. Balatsoukas-Stimming, "Joint detection and self-interference cancellation in full-duplex systems using machine learning," in *Proc. 55th Asilomar Conf. Signals, Syst., Comput.*, Mar. 2021, pp. 989–992.

[2] A. Balatsoukas-Stimming, "End-to-end learned self-interference cancellation," in *Proc. 56th Asilomar Conf. Signals, Syst., Comput.*, Mar. 2022, pp. 1334–1338.

[3] D. H. Kong, Y.-S. Kil, and S.-H. Kim, "Neural network aided digital self-interference cancellation for full-duplex communication over time-varying channels," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6201–6213, Jun. 2022.

[4] O. Ploder *et al.*, "SICNet—Low complexity sample adaptive neural network-based self-interference cancellation in LTE-A/5G mobile transceivers," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 958–972, Jun. 2022.

[5] Y. Chen *et. al.*, "MIMO full duplex radios with deep learning," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jul. 2020, pp. 1–6.

References

Chapter 1

[1] L. Bariah, L. Mohjazi, S. Muhaidat, P. C. Sofotasios, G. K. Kurt, H. Yanikomeroglu, and O. A. Dobre, "A prospective look: Key enabling technologies, applications and open research topics in 6G networks," *IEEE Access*, vol. 8, pp. 174792–174820, Aug. 2020.

[2] P. Yang, Y. Xiao, M. Xiao, and S. Li, "6G wireless communications: Vision and potential techniques," *IEEE Netw.*, vol. 33, no. 4, pp. 70–75, Jul. 2019.

[3] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?," *Nature Electron.*, vol. 3, no. 1, pp. 20–29, Jan. 2020.

[4] X. You *et al.*, "Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Sci. China Inf. Sci.*, vol. 64, no. 1, pp. 1–74, Nov. 2020.

[5] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, D. Niyato, O. A. Dobre, H. V. Poor, "6G internet of things: A comprehensive survey," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 359–383, Jan. 2022.

[6] F. Guo, F. R. Yu, H. Zhang, X. Li, H. Ji, and V. C. M. Leung, "Enabling massive IoT toward 6G: A comprehensive survey," *IEEE Internet Things*

J., vol. 8, no. 15, pp. 11891–11915, Aug. 2021.

[7] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

[8] 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.

[9] 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.

[10] B. Debaillie *et al.*, "Analog/RF solutions enabling compact full-duplex radios," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1662–1673, Sep. 2014.

[11] Y.-S. Choi and H. Shirani-Mehr, "Simultaneous transmission and reception: Algorithm, design and system level performance," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 12, pp. 5992–6010, Dec. 2013.

[12] L. Laughlin, M. A. Beach, K. A. Morris, and J. L. Haine, "Optimum single antenna full duplex using hybrid junctions," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1653–1661, Sep. 2014.

[13] D. Korpi, L. Anttila, V. Syrjala, and M. Valkama, "Widely linear digital self-interference cancellation in direct-conversion fullduplex transceiver," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1674–1687, Sep. 2014.

[14] D. Korpi, *et al*, "Digital self-interference cancellation under nonideal RF components: Advanced algorithms and measured performance," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2015, pp. 286–290.

[15] M. A. Tafreshi, M. Koskela, D. Korpi, P. Jääskeläinen, M. Valkama, and J. Takala, "Software defined radio implementation of adaptive nonlinear digital self-interference cancellation for mobile inband fullduplex radio," in *Proc. IEEE Global Conf. Signal Inform. Process. (GLOBALSIP)*, Dec. 2016, pp. 733–737.

[16] M. S. Amjad and O. Gurbuz, "Linear digital cancellation with reduced computational complexity for full-duplex radios," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[17] D. Korpi, L. Anttila, and M. Valkama, "Nonlinear self-interference cancellation in MIMO full-duplex transceivers under crosstalk," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, pp. 1–15, Dec. 2017.

[18] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[19] E. Nachmani, E. Marciano, L. Lugosch, W. J. Gross, D. Burshtein, and Y. Be'ery, "Deep learning methods for improved decoding of linear codes," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 119–131, Feb. 2018.

[20] F. Liang, C. Shen, and F. Wu, "An iterative BP-CNN architecture for channel decoding," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 144–159, Feb. 2018.

[21] M. Gao, Y. Li, O. A. Dobre, and N. Al-Dhahir, "Joint blind identification of the number of transmit antennas and MIMO schemes using Gerschgorin radii and FNN," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 373–387, Jan. 2019.

[22] X. Cheng, D. Liu, C. Wang, S. Yan, and Z. Zhu, "Deep learning-based

channel estimation and equalization scheme for fbmc/oqam systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 881–884, Jun. 2019.

[23] M. Xu, S. Zhang, J. Ma, and O. A. Dobre, "Deep learning-based time-varying channel estimation for RIS assisted communication," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 94–98, Jan. 2022.

[24] M. Al-Nahhal, I. Al-Nahhal, O. A. Dobre, X. Lin, D. Chang, and C. Li, "Joint estimation of linear and nonlinear coherent optical fiber signal to-noise ratio," *IEEE Photon. Technol. Lett.*, vol. 35, no. 1, pp. 23–26, Nov. 2022.

[25] M. Al-Nahhal, I. Al-Nahhal, O. A. Dobre, S. K. O. Soman, D. Chang, and C. Li, "Learned signal-to-noise ratio estimation in optical fiber communication links," *IEEE Photon. J.*, vol. 14, no. 6, pp. 1–7, Nov. 2022.

[26] Y. Liu, I. Al-Nahhal, O. A. Dobre, and F. Wang, "Deep-learning-based channel estimation for IRS-assisted ISAC system," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2022, pp. 4220–4225.

[27] Y. Liu, I. Al-Nahhal, O. A. Dobre, and F. Wang, "Deep-learning channel estimation for IRS-assisted integrated sensing and communication system," *IEEE Trans. Veh Technol.*, early access, pp. 1–14, Dec. 2022.

[28] E. A. Makled, I. Al-Nahhal, O. A. Dobre, and O. Üreten, "Identification of cellular signal measurements using machine learning," *IEEE Trans. Instrum. Meas.*, vol. 72, no. 1, pp. 1–4, Jan. 2023.

[29] 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.

[30] 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.

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

[32] 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. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 589–593.

[33] 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.

[34] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Advanced machine learning techniques for self-interference cancellation in full-duplex radios," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 1149–1153.

[35] M. Erdem, H. Ozkan, and O. Gurbuz, "Nonlinear digital self-interference cancellation with SVR for full duplex communication," in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC)*, Jun. 2020, pp. 1–6.

[36] M. Yilan, O. Gurbuz, and H. Ozkan, "Integrated linear and nonlinear digital cancellation for full duplex communication," *IEEE Wireless Communications*, vol. 28, no. 1, pp. 20–27, Feb. 2021.

[37] F. Jochems and A. Balatsoukas-Stimming, "Non-linear self-interference cancellation via tensor completion," in *Proc. 54th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2020, pp. 905–909.

[38] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Identification of non-linear RF systems using backpropagation," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2020, pp. 1–6.

[39] M. Erdem, H. Ozkan, and O. Gurbuz, "A new online nonlinear self-interference cancelation method with random Fourier features," *IEEE Wireless Commun. Lett.*, vol. 11, no. 7, pp. 1379–1383, Apr. 2022.

[40] J. Chen, L. Zhang, and Y.-C. Liang, "Exploiting Gaussian mixture model clustering for full-duplex transceiver design," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5802–5816, Aug. 2019.

[41] O. Zhao, W-S Liao, K. Li, T. Matsumura, F. Kojima, and H. Harada, "Lazy learning-based self-interference cancellation for wireless communication systems with in-band full-duplex operations," in *Proc. IEEE 32nd Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2021, pp. 1589–1594.

[42] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Low complexity neural network structures for self-interference cancellation in full-duplex radio," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 181–185, Jan. 2021.

[43] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Full-duplex self-interference cancellation using dual-neurons neural networks," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 557–561, Mar. 2022.

[44] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Hybrid-layers neural network architectures for modeling the self-interference in full-duplex systems," *IEEE Trans. Veh Technol.*, vol. 71, no. 6, pp. 6291–6307, Jun. 2022.

[45] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang,

"Residual neural networks for learning the full-duplex self-interference," in *Proc. 57th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2023, pp. 1–4.

[46] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Machine learning-based self-interference cancellation for full-duplex radio: Approaches, open challenges, and future research directions," *IEEE Open J. Veh. Technol.*, vol. 5, no. 1, pp. 21–47, Nov. 2023.

[47] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Extreme learning machine-assisted full-duplex self-interference cancellation," *to be submitted to IEEE Trans. Instrum. Meas.*, May 2024.

Chapter 2

[1] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, Mar. 2020.

[2] 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.

[3] H. V. Nguyen et al., "On the spectral and energy efficiencies of fullduplex cell-free massive MIMO," *IEEE J. Sel. Areas Commun.,* vol. 38, no. 8, pp. 1698–1718, Aug. 2020.

[4] F. Chen, R. Morawski, and T. Le-Ngoc, "Self-interference channel characterization for wideband 2 × 2 MIMO full-duplex transceivers using dual-polarized antennas," *IEEE Trans. Antennas Propag.*, vol. 66, no. 4, pp. 1967–1976, Apr. 2018.

[5] G. Agrawal, S. Aniruddhan, and R. K. Ganti, "A compact mixer-first receiver with >24 dB self-interference cancellation for fullduplex radios," *IEEE Microw. Wireless Compon. Lett.*, vol. 26, no. 12, pp. 1005–1007, Dec. 2016.

[6] Z. Zhang, K. Long, A. V. Vasilakos, and L. Hanzo, "Full-duplex wireless communications: Challenges, solutions, and future research directions," *Proc. IEEE*, vol. 104, no. 7, pp. 1369–1409, Jul. 2016.

[7] D. Korpi, L. Anttila, and M. Valkama, "Nonlinear self-interference cancellation in MIMO full-duplex transceivers under crosstalk," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, pp. 1–15, Dec. 2017.

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

[9] Y. Yang, S. Zhang, F. Gao, J. Ma, and O. A. Dobre, "Graph neural network based channel tracking for massive MIMO networks," Apr. 2020. [Online]. Available: https://arxiv.org/abs/2004.08738v1.

[10] 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.

[11] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Advanced machine learning techniques for self-interference cancellation in full-duplex radios," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 1149–1153.

[12] B. Warsito, R. Santoso, and H. Yasin, "Cascade forward neural network

for time series prediction," *J. Phys. Conf. Ser.*, vol. 1025, no. 1, pp. 1-9, May 2018.

[13] S. E. Fahlman and C. Lebiere, "The cascade-correlation learning architecture," in *Proc. Adv. Neural Inf. Process. Syst.*, Jun. 1990, pp. 524–532.

[14] F. J. Śmieja, "Neural network constructive algorithms: Trading generalization for learning efficiency?," *Circuits Syst., Signal Processing*, vol. 12, no. 2, pp. 331–374, Jun. 1993.

[15] T. Y. Kwok and D. Y. Yeung, "Constructive algorithms for structure learning in feedforward neural networks for regression problems," *IEEE Trans. Neural Networks*, vol. 8, no. 3, pp. 630–645, May 1997.

[16] D. Ruta, B. Gabrys, and C. Lemke, "A generic multilevel architecture for time series prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 3, pp. 350–359, Mar. 2011.

Chapter 3

[1] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, Mar. 2020.

[2] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May 2020.

[3] H. V. Nguyen, V. Nguyen, O. A. Dobre, S. K. Sharma, S. Chatzinotas, B. Ottersten, and O. Shin, "On the spectral and energy efficiencies of full-duplex cell-free massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1698–1718, Jun. 2020.

[4] 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, Sept. 2014.

[5] D. Kim, H. Lee, and D. Hong, "A survey of in-band full-duplex transmission: From the perspective of PHY and MAC layers," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2017–2046, Feb. 2015.

[6] Z. Zhang, K. Long, A. V. Vasilakos, and L. Hanzo, "Full-duplex wireless communications: Challenges, solutions, and future research directions," *Proc. IEEE*, vol. 104, no. 7, pp. 1369–1409, Jul. 2016.

[7] 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.

[8] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

[9] E. Everett, A. Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 680–694, Feb. 2014.

[10] 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. Comput.*, Nov. 2013, pp. 992–997.

[11] M. Duarte and A. Sabharwal, "Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results," in *Proc. 44th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2010, pp. 1558–1562.

[12] Y.-S. Choi and H. Shirani-Mehr, "Simultaneous transmission and re-

ception: Algorithm, design and system level performance," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 12, pp. 5992–6010, Dec. 2013.

[13] B. Debaillie et al., "Analog/RF solutions enabling compact full-duplex radios," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1662–1673, Sep. 2014.

[14] L. Laughlin, M. A. Beach, K. A. Morris, and J. L. Haine, "Optimum single antenna full duplex using hybrid junctions," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1653–1661, Sep. 2014.

[15] E. Ahmed and A. M. Eltawil, "All-digital self-interference cancellation technique for full-duplex systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3519–3532, Jul. 2015.

[16] D. Korpi, L. Anttila, and M. Valkama, "Nonlinear self-interference cancellation in MIMO full-duplex transceivers under crosstalk," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, pp. 1–15, Dec. 2017.

[17] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[18] M. Gao, Y. Li, O. A. Dobre, and N. Al-Dhahir, "Joint blind identification of the number of transmit antennas and MIMO schemes using Gerschgorin radii and FNN," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 373–387, Jan. 2019.

[19] H. Ye, G. Y. Li, and B.-H. F. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018.

[20] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless*

Commun. Lett., vol. 7, no. 5, pp. 852–855, Oct. 2018.

[21] X. Cheng, D. Liu, C. Wang, S. Yan, and Z. Zhu, "Deep learning-based channel estimation and equalization scheme for fbmc/oqam systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 881–884, Jun. 2019.

[22] T. Gruber, S. Cammerer, J. Hoydis, and S. T. Brink, "On deep learning-based channel decoding," in *Proc. IEEE 51st Annu. Conf. Inf. Sci. Syst. (CISS)*, Baltimore, MD, USA, Mar. 2017, pp. 1–6.

[23] T. Liu, S. Boumaiza, and F. M. Ghannouchi, "Dynamic behavioral modeling of 3G power amplifiers using real-valued time delay neural networks," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 3, pp. 1025– 1033, Mar. 2004.

[24] F. Mkadem and S. Boumaiza, "Physically inspired neural network model for RF power amplifier behavioral modeling and digital predistortion," *IEEE Trans. Microw. Theory Technol.*, vol. 59, no. 4, pp. 913–923, Apr. 2011.

[25] 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.

[26] C. Tarver, L. Jiang, A. Sefidi, and J. Cavallaro, "Neural network DPD via backpropagation through a neural network model of the PA," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 358–362.

[27] 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.

[28] C. Häger and H. D. Pfister, "Nonlinear interference mitigation via deep neural networks," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, San

Diego, CA, USA, Mar. 2018, pp. 1–3.

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

[30] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Advanced machine learning techniques for self-interference cancellation in full-duplex radios," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 1149–1153.

[31] 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. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 589–593.

[32] 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.

[33] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Low complexity neural network structures for self-interference cancellation in full-duplex radio," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 181-185, Jan. 2021.

[34] M. Isaksson, D. Wisell, and D. Ronnow, "A comparative analysis of behavioral models for RF power amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 1, pp. 348–359, Jan. 2006.

[35] R. Zayani, R. Bouallegue, and D. Roviras, "Adaptive predistortions based on neural networks associated with levenberg-marquardt algorithm for

satellite down links," *EURASIP J. Wireless Commun. Net.,* vol. 2008, no. 1, p. 2, Jul. 2008.

[36] J. X. Peng, K. Li, and D. S. Huang, "A hybrid forward algorithm for RBF neural network construction," *IEEE Trans. Neural Netw.*, vol. 17, no. 6, pp. 1439–1451, Nov. 2006.

[37] X. Xia, K. Xu, D. Zhang, Y. Xu, and Y. Wang, "Beam-domain full-duplex massive MIMO: Realizing co-time co-frequency uplink and downlink transmission in the cellular system," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 8845–8862, Oct. 2017.

[38] K. Xu, Z. Shen, Y. Wang, X. Xia, and D. Zhang, "Hybrid time-switching and power splitting SWIPT for full-duplex massive MIMO systems: A beam-domain approach," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7257–7274, Aug. 2018.

[39] X. Xia, K. Xu, Y. Wang, and Y. Xu, "A 5G-enabling technology: Benefits, feasibility, and limitations of in-band full-duplex mmimo," *IEEE Veh. Technol. Mag.*, vol. 13, no. 3, pp. 81–90, Sep. 2018.

[40] Y. Chen et. al, "MIMO full duplex radios with deep learning," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2020, pp. 1–6.

[41] C. Auer, K. Kostoglou, T. Paireder, O. Ploder, and M. Huemer, "Support vector machines for self-interference cancellation in mobile communication transceivers," in *Proc. IEEE Veh. Technol. Conf. (VTC-Spring)*, 2020, pp. 1–6.

Chapter 4

[1] A. T. Le *et* al., "Analog least mean square adaptive filtering for self-interference cancellation in full duplex radios," *IEEE Wireless Commun.*, vol. 28, no. 1, pp. 12–18, Feb. 2021.

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

[3] D. Korpi, L. Anttila, and M. Valkama, "Nonlinear self-interference cancellation in MIMO full-duplex transceivers under crosstalk," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, pp. 1–15, Dec. 2017.

[4] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Advanced machine learning techniques for self-interference cancellation in full-duplex radios," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 1149–1153.

[5] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Low complexity neural network structures for self-interference cancellation in full-duplex radio," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 181-185, Jan. 2021.

[6] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Hybrid-layers neural network architectures for modeling the self-interference in full-duplex systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6291–6307, Jun. 2022.

[7] F. Mkadem *et* al., "Behavioral modeling and digital predistortion of power amplifiers with memory using two hidden layers artificial neural networks," in *Proc. IEEE MTT-S Int. Microw. Symp.*, 2010, pp. 656–659.

[8] F. Mkadem and S. Boumaiza, "Physically inspired neural network model for RF power amplifier behavioral modeling and digital predistortion," *IEEE Trans. Microw. Theory Technol.*, vol. 59, no. 4, pp. 913–923, Apr. 2011.

[9] T. Liu *et* al., "Dynamic behavioral modeling of 3G power amplifiers using real-valued time delay neural networks," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 3, pp. 1025–1033, Mar. 2004.

[10] Y. Kurzo *et* al., "Hardware implementation of neural self-interference cancellation," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 10, no. 2, pp. 204-216, Jun. 2020.

Chapter 5

[1] L. Bariah, L. Mohjazi, S. Muhaidat, P. C. Sofotasios, G. K. Kurt, H. Yanikomeroglu, and O. A. Dobre, "A prospective look: Key enabling technologies, applications and open research topics in 6G networks," *IEEE Access*, vol. 8, pp. 174792–174820, Aug. 2020.

[2] P. Yang, Y. Xiao, M. Xiao, and S. Li, "6G wireless communications: Vision and potential techniques," *IEEE Netw.*, vol. 33, no. 4, pp. 70–75, Jul. 2019.

[3] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?," *Nature Electron.*, vol. 3, no. 1, pp. 20–29, Jan. 2020.

[4] X. You *et al.*, "Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Sci. China Inf. Sci.*, vol. 64, no. 1, pp. 1–74, Nov. 2020.

[5] Z. Zhang *et al.*, "6G wireless networks: Vision, requirements, architecture,

and key technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 28–41, Sep. 2019.

[6] E. C. Strinati *et al.*, "6G: The next frontier: From holographic messaging to artificial intelligence using subterahertz and visible light communication," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 42–50, Sep. 2019.

[7] K. David and H. Berndt, "6G vision and requirements: Is there any need for beyond 5G?" *IEEE Veh. Technol. Mag.*, vol. 13, no. 3, pp. 72–80, Sep. 2018.

[8] S. Zhang, C. Xiang, and S. Xu, "6G: Connecting everything by 1000 times price reduction," *IEEE Open J. Veh. Technol.*, vol. 1, pp. 107–115, Mar. 2020.

[9] H. Viswanathan and P. E. Mogensen, "Communications in the 6G era," *IEEE Access*, vol. 8, pp. 57063–57074, Mar. 2020.

[10] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, Mar. 2020.

[11] S. Chen, Y.-C. Liang, S. Sun, S. Kang, W. Cheng, and M. Peng, "Vision, requirements, and technology trend of 6G: How to tackle the challenges of system coverage, capacity, user data-rate and movement speed," *IEEE Wireless Commun. Mag.*, vol. 27, no. 2, pp. 218–228, Apr. 2020.

[12] F. Tariq, M. R. A. Khandaker, K.-K. Wong, M. A. Imran, M. Bennis, and M. Debbah, "A speculative study on 6G," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 118–125, Aug. 2020.

[13] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions," *IEEE Open J. Commun. Soc.*, vol. 1,

pp. 957–975, Jul. 2020.

[14] I. F. Akyildiz, A. Kak, and S. Nie, "6G and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133995–134030, July 2020.

[15] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May 2020.

[16] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, Aug. 2019.

[17] B. Zong, C. Fan, X. Wang, X. Duan, B. Wang, and J. Wang, "6G technologies: Key drivers, core requirements, system architectures, and enabling technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 18–27, Sep. 2019.

[18] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 334–366, Feb. 2021.

[19] A. Shahraki, M. Abbasi, M. J. Piran, and A. Taherkordi, "A comprehensive survey on 6G networks: Applications, core services, enabling technologies, and future challenges," Jan. 2021. [Online]. Available: arXiv:2101.12475.

[20] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, D. Niyato, O. A. Dobre, H. V. Poor, "6G internet of things: A comprehensive survey," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 359–383, Jan. 2022.

[21] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*,

vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

[22] H. V. Nguyen, V.-D. Nguyen, O. A. Dobre, and O.-S. Shin, "Sum rate maximization based on sub-array antenna selection in a full-duplex system," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.

[23] A. Yadav, O. A. Dobre, and N. Ansari, "Energy and traffic aware full-duplex communications for 5G systems," *IEEE Access*, vol. 5, pp. 11278–11290, May 2017.

[24] A. Yadav, O. A. Dobre, and H. V. Poor, "Is self-interference in full-duplex communications a foe or a friend?" *IEEE Signal Process. Lett.*, vol. 25, no. 7, pp. 951–955, Jul. 2018.

[25] E. Everett, A. Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 680–694, Jan. 2014.

[26] S. Hong *et al.*, "Applications of self-interference cancellation in 5G and beyond," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 114–121, Feb. 2014.

[27] D. Korpi *et al.*, "Full-duplex transceiver system calculations: Analysis of ADC and linearity challenges," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3821–3836, Jul. 2014.

[28] 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.

[29] A. Yadav, G. I. Tsiropoulos, and O. A. Dobre, "Full-duplex communications: Performance in ultradense mm-wave small-cell wireless networks," *IEEE Veh. Technol. Mag.*, vol. 13, no. 2, pp. 40–47, Jun. 2018.

[30] Z. Zhang, X. Chai, K. Long, A. V. Vasilakos, and L. Hanzo, "Full duplex

techniques for 5G networks: Self-interference cancellation, protocol design, and relay selection," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 128–137, May 2015.

[31] M. Chung, M. S. Sim, J. Kim, D. K. Kim, and C.-B. Chae, "Prototyping real-time full duplex radios," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 56–63, Sep. 2015.

[32] D. Kim, H. Lee, and D. Hong, "A survey of in-band full-duplex transmission: From the perspective of PHY and MAC layers," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2017–2046, 4th quart., 2015.

[33] Z. Zhang, K. Long, A. V. Vasilakos, and L. Hanzo, "Full-duplex wireless communications: Challenges, solutions, and future research directions," *Proc. IEEE*, vol. 104, no. 7, pp. 1369–1409, Jul. 2016.

[34] D. Korpi *et al.*, "Full-duplex mobile device: Pushing the limits," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 80–87, Sep. 2016.

[35] M. Amjad, F. Akhtar, M. H. Rehmani, M. Reisslein, and T. Umer, "Full-duplex communication in cognitive radio networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2158–2191, 4th Quart., 2017.

[36] V.-D. Nguyen, H. V. Nguyen, O. A. Dobre, and O.-S. Shin, "A new design paradigm for secure full-duplex multiuser systems," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 7, pp. 1480–1498, Jul. 2018.

[37] H. V. Nguyen, V.-D. Nguyen, O. A. Dobre, Y. Wu, and O.-S. Shin, "Joint antenna array mode selection and user assignment for full-duplex MU-MISO systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 2946–2963, Jun. 2019.

[38] H. V. Nguyen, V.-D. Nguyen, O. A. Dobre, D. N. Nguyen, E. Dutkiewicz, and O.-S. Shin, "Joint power control and user association for NOMA-

based full-duplex systems," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 8037–8055, Nov. 2019.

[39] Q. N. Le, N.-P. Nguyen, A. Yadav, and O. A. Dobre, "Outage performance of full-duplex overlay CR-NOMA networks with SWIPT," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[40] H. V. Nguyen, V.-D. Nguyen, O. A. Dobre, N.-P. Nguyen, R. Zhao, and S. Chatzinotas, "On the spectral and energy efficiencies of full-duplex cell-free massive MIMO," *IEEE J. Select. Areas Commun.*, vol. 38, no. 8, pp. 1698–1718, Aug. 2020.

[41] Q. N. Le, A. Yadav, N.-P. Nguyen, O. A. Dobre, and R. Zhao, "Full-duplex non-orthogonal access cooperative overlay spectrum-sharing networks with SWIPT," *IEEE Trans. Green Commun. Netw.*, vol. 5, no. 1, pp. 322–334, Mar. 2021.

[42] E. A. Makled and O. A. Dobre, "On the security of full-duplex relay-assisted underwater acoustic network with NOMA," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6255–6265, Jun. 2022.

[43] 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.

[44] B. Debaillie *et al.*, "Analog/RF solutions enabling compact full-duplex radios," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1662–1673, Sep. 2014.

[45] Y.-S. Choi and H. Shirani-Mehr, "Simultaneous transmission and reception: Algorithm, design and system level performance," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 12, pp. 5992–6010, Dec. 2013.

[46] L. Laughlin, M. A. Beach, K. A. Morris, and J. L. Haine, "Optimum

single antenna full duplex using hybrid junctions," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1653–1661, Sep. 2014.

[47] D. Korpi, L. Anttila, V. Syrjala, and M. Valkama, "Widely linear digital self-interference cancellation in direct-conversion fullduplex transceiver," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1674–1687, Sep. 2014.

[48] D. Korpi, L. Anttila, and M. Valkama, "Reference receiver based digital self-interference cancellation in MIMO full-duplex transceivers," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 1001–1007.

[49] D. Korpi, *et al*, "Digital self-interference cancellation under nonideal RF components: Advanced algorithms and measured performance," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2015, pp. 286–290.

[50] E. Ahmed and A. M. Eltawil, "All-digital self-interference cancellation technique for full-duplex systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3519–3532, Jul. 2015.

[51] D. Korpi, Y.-S. Choi, T. Huusari, L. Anttila, S. Talwar, and M. Valkama, "Adaptive nonlinear digital self-interference cancellation for mobile inband full-duplex radio: Algorithms and RF measurements," in *Proc. IEEE IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–7.

[52] M. A. Tafreshi, M. Koskela, D. Korpi, P. Jääskeläinen, M. Valkama, and J. Takala, "Software defined radio implementation of adaptive nonlinear digital self-interference cancellation for mobile inband fullduplex radio," in *Proc. IEEE Global Conf. Signal Inform. Process. (GLOBALSIP)*, Dec. 2016, pp. 733–737.

[53] M. S. Amjad and O. Gurbuz, "Linear digital cancellation with reduced computational complexity for full-duplex radios," in *Proc. IEEE Wireless*

Commun. Netw. Conf. (WCNC), Mar. 2017, pp. 1–6.

[54] X. Quan, Y. Liu, S. Shao, C. Huang, and Y. Tang, "Impacts of phase noise on digital self-interference cancellation in full-duplex communications," *IEEE Trans. Signal Process.*, vol. 65, no. 7, pp. 1881–1893, Apr. 2017.

[55] Y. Liu, X. Quan, W. Pan, and Y. Tang, "Digitally assisted analog interference cancellation for in-band full-duplex radios," *IEEE Commun. Lett.*, vol. 21, no. 5, pp. 1079–1082, May 2017.

[56] P. Ferrand and M. Duarte, "Multi-tap digital canceller for full-duplex applications," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2017, pp. 1–5.

[57] D. Korpi, L. Anttila, and M. Valkama, "Nonlinear self-interference cancellation in MIMO full-duplex transceivers under crosstalk," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, pp. 1–15, Dec. 2017.

[58] R. V. Kulkarni, A. Forster, and G. K. Venayagamoorthy, "Computational intelligence in wireless sensor networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 1, pp. 68–96, 1st Quart., 2011.

[59] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine-learning techniques in cognitive radios," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1136–1159, 3rd Quart., 2013.

[60] M. A. Alsheikh, S. Lin, D. Niyato, and H.-P. Tan, "Machine learning in wireless sensor networks: Algorithms, strategies, and applications," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1996–2018, 4th Quart., 2014.

[61] H. A. Al-Rawi, M. A. Ng, and K.-L. A. Yau, "Application of reinforcement learning to routing in distributed wireless networks: A review," *Artif. Intell. Rev.*, vol. 43, no. 3, pp. 381–416, Jan. 2015.

[62] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A survey of

machine learning techniques applied to self-organizing cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2392–2431, 4th Quart., 2017.

[63] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[64] E. Nachmani, E. Marciano, L. Lugosch, W. J. Gross, D. Burshtein, and Y. Be'ery, "Deep learning methods for improved decoding of linear codes," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 119–131, Feb. 2018.

[65] F. Liang, C. Shen, and F. Wu, "An iterative BP-CNN architecture for channel decoding," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 144–159, Feb. 2018.

[66] F. Pacheco, E. Exposito, M. Gineste, C. Baudoin, and J. Aguilar, "Towards the deployment of machine learning solutions in network traffic classification: A systematic survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1988–2014, 2nd Quart., 2018.

[67] D. He, C. Liu, T. Q. S. Quek, and H. Wang, "Transmit antenna selection in MIMO wiretap channels: A machine learning approach," *IEEE Wireless Commun. Lett.*, vol. 7, no. 4, pp. 634–637, Aug. 2018.

[68] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 852–855, Oct. 2018.

[69] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2595–2621, 4th Quart., 2018.

[70] M. Gao, Y. Li, O. A. Dobre, and N. Al-Dhahir, "Joint blind identification of the number of transmit antennas and MIMO schemes using Gerschgorin radii and FNN," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 373–387, Jan. 2019.

[71] J. Xie *et al.*, "A survey of machine learning techniques applied to software defined networking (SDN): Research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 393–430, 1st Quart., 2019.

[72] M. Usama *et al.*, "Unsupervised machine learning for networking: Techniques, applications and research challenges," *IEEE Access*, vol. 7, pp. 65579–65615, May 2019.

[73] X. Cheng, D. Liu, C. Wang, S. Yan, and Z. Zhu, "Deep learning-based channel estimation and equalization scheme for fbmc/oqam systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 881–884, Jun. 2019.

[74] M. Chen, *et al.*, "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3039–3071, 4th Quart., 2019.

[75] N. C. Luong *et al.*, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3133–3174, 4th Quart., 2019.

[76] Y. Sun, M. Peng, Y. Zhou, Y. Huang, and S. Mao, "Application of machine learning in wireless networks: Key techniques and open issues," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3072–3108, 4th Quart., 2019.

[77] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation, opportunities and challenges," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 46–51, Jun. 2020.

[78] A. Faisal, I. Al-Nahhal, O. A. Dobre, and T. M. N. Ngatched, "Deep rein-

forcement learning for optimizing RIS-assisted HD-FD wireless systems," *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3893–3897, Dec. 2021.

[79] S. Zhang, S. Zhang, F. Gao, J. Ma, and O. A. Dobre, "Deep learning based RIS channel extrapolation with element-grouping," *IEEE Wireless Commun. Lett.*, vol. 10, no. 12, pp. 2644–2648, Dec. 2021.

[80] M. Xu, S. Zhang, J. Ma, and O. A. Dobre, "Deep learning-based time-varying channel estimation for RIS assisted communication," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 94–98, Jan. 2022.

[81] A. Faisal, I. Al-Nahhal, O. A. Dobre, and T. M. N. Ngatched, "Deep reinforcement learning for RIS-assisted FD systems: Single or distributed RIS?," *IEEE Commun. Lett.*, vol. 26, no. 7, pp. 1563–1567, Jul. 2022.

[82] M. Al-Nahhal, I. Al-Nahhal, O. A. Dobre, X. Lin, D. Chang, and C. Li, "Joint estimation of linear and nonlinear coherent optical fiber signal to-noise ratio," *IEEE Photon. Technol. Lett.*, vol. 35, no. 1, pp. 23–26, Nov. 2022.

[83] M. Al-Nahhal, I. Al-Nahhal, O. A. Dobre, S. K. O. Soman, D. Chang, and C. Li, "Learned signal-to-noise ratio estimation in optical fiber communication links," *IEEE Photon. J.*, vol. 14, no. 6, pp. 1–7, Nov. 2022.

[84] A. Faisal, I. Al-Nahhal, O. A. Dobre, and T. M. N. Ngatched, "Distributed RIS-assisted FD systems with discrete phase shifts: A reinforcement learning approach," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2022, pp. 5862–5867.

[85] Y. Liu, I. Al-Nahhal, O. A. Dobre, and F. Wang, "Deep-learning-based channel estimation for IRS-assisted ISAC system," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2022, pp. 4220–4225.

[86] Y. Liu, I. Al-Nahhal, O. A. Dobre, and F. Wang, "Deep-learning channel

estimation for IRS-assisted integrated sensing and communication system," *IEEE Trans. Veh Technol.*, early access, pp. 1–14, Dec. 2022.

[87] E. A. Makled, I. Al-Nahhal, O. A. Dobre, and O. Üreten, "Identification of cellular signal measurements using machine learning," *IEEE Trans. Instrum. Meas.*, vol. 72, no. 1, pp. 1–4, Jan. 2023.

[88] 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.

[89] 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.

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

[91] 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.*, Dec. 2019, pp. 53–57.

[92] K. E. Kolodziej, A. U. Cookson, and B. T. Perry, "Neural network tuning for analog-RF self-interference cancellation," in *Proc. IEEE MTT-S Int. Microw. Symp.*, June 2021, pp. 673–676.

[93] K. E. Kolodziej, A. U. Cookson, and B. T. Perry, "RF canceller tuning acceleration using neural network machine learning for in-band full-duplex systems," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1330–1334, May 2021.

[94] V. Tapio and M. Juntti, "Non-linear self-interference cancelation for full-duplex transceivers based on Hammerstein-Wiener model," *IEEE Com-*

mun. Lett., vol. 25, no. 11, pp. 3684–3688, Nov. 2021.

[95] 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. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 589–593.

[96] 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.

[97] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Advanced machine learning techniques for self-interference cancellation in full-duplex radios," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 1149–1153.

[98] Q. Wang, F. He, and J. Meng, "Performance comparison of real and complex valued neural networks for digital self-interference cancellation," in *Proc. IEEE 19th Int. Conf. Commun. Technol. (ICCT)*, Oct. 2019, pp. 1193–1199.

[99] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Low complexity neural network structures for self-interference cancellation in full-duplex radio," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 181–185, Jan. 2021.

[100] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Hybrid-layers neural network architectures for modeling the self-interference in full-duplex systems," *IEEE Trans. Veh Technol.*, vol. 71, no. 6, pp. 6291–6307, Jun. 2022.

[101] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang,

"Full-duplex self-interference cancellation using dual-neurons neural networks," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 557–561, Mar. 2022.

[102] D. H. Kong, Y.-S. Kil, and S.-H. Kim, "Neural network aided digital self-interference cancellation for full-duplex communication over time-varying channels," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6201–6213, Jun. 2022.

[103] 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 Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[104] W. Zhang, J. Yin, D. Wu, G. Guo, and Z. Lai, "A self-interference cancellation method based on deep learning for beyond 5G full-duplex system," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, Sep. 2018, pp. 1–5.

[105] K. Muranov, M. A. Islam, B. Smida, and N. Devroye, "On deep learning assisted self-interference estimation in a full-duplex relay link," *IEEE Wireless Commun. Lett.*, vol. 10, no. 12, pp. 2762–2766, Dec. 2021.

[106] C. Auer, K. Kostoglou, T. Paireder, O. Ploder, and M. Huemer, "Support vector machines for self-interference cancellation in mobile communication transceivers," in *Proc. IEEE Veh. Technol. Conf. (VTC-Spring)*, Jun. 2020, pp. 1–6.

[107] M. Erdem, H. Ozkan, and O. Gurbuz, "Nonlinear digital self-interference cancellation with SVR for full duplex communication," in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC)*, Jun. 2020, pp. 1–6.

[108] M. Yilan, O. Gurbuz, and H. Ozkan, "Integrated linear and nonlinear digital cancellation for full duplex communication," *IEEE Wireless Communications*, vol. 28, no. 1, pp. 20–27, Feb. 2021.

[109] F. Jochems and A. Balatsoukas-Stimming, "Non-linear self-interference cancellation via tensor completion," in *Proc. 54th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2020, pp. 905–909.

[110] 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. (ICNC)*, Mar. 2018, pp. 779–783.

[111] K. Wongsuphasawat, D. Smilkov, J. Wexler, J. Wilson, D. Mane, D. Fritz, D. Krishnan, F. B. Viegas, and M. Wattenberg, "Visualizing dataflow graphs of deep learning models in TensorFlow," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 1, pp. 1–12, Jan. 2018.

[112] M. Erdem, H. Ozkan, and O. Gurbuz, "A new online nonlinear self-interference cancelation method with random Fourier features," *IEEE Wireless Commun. Lett.*, vol. 11, no. 7, pp. 1379–1383, Apr. 2022.

[113] R. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Proc. 20th Int. Conf. Adv. Neural Inf. Process. Syst.*, Dec. 2007, pp. 1177–1184.

[114] J. Lu, S. C. Hoi, J. Wang, P. Zhao, and Z.-Y. Liu, "Large scale online kernel learning," *J. Mach. Learn. Res.*, vol. 17, no. 47, pp. 1–43, Jan. 2016.

[115] S. Mehrkanoon and J. A. K. Suykens, "Deep hybrid neural-kernel networks using random Fourier features," *Neurocomputing*, vol. 298, no. 7, pp. 46–54, Jul. 2018.

[116] R. Mitra, V. Bhatia, S. Jain, and K. Choi, "Performance analysis of random Fourier features based unsupervised multistage-clustering for VLC," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2659–2663, Aug. 2021.

[117] R. Mitra, *et al.*, "Analytical guarantees for hyperparameter free RFF based deep learning in the low-data regime," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 2, pp. 634–638, Feb. 2022.

[118] R. Mitra and G. Kaddoum, "Random Fourier feature based deep learning for wireless communications," *IEEE Trans. Cogn. Commun. Netw.*, vol. 11, no. 2, pp. 468–479., Jun. 2022.

[119] Y.-H. Lin, Y.-T. Liao, J.-Y. Chu, P.-J. Su, and T.-Y. Hsu, "Digital self interference cancellation via dynamic regression for in-band full-duplex system," in *Proc. IEEE 5th Global Conf. Consum. Electron.*, Oct. 2016, pp. 1–3.

[120] J. Chen, L. Zhang, and Y.-C. Liang, "Exploiting Gaussian mixture model clustering for full-duplex transceiver design," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5802–5816, Aug. 2019.

[121] X. F. He, D. Cai, Y. L. Shao, H. J. Bao, and J. W. Han, "Laplacian regularized Gaussian mixture model for data clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 9, pp. 1406–1418, Sep. 2011.

[122] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Identification of non-linear RF systems using backpropagation," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2020, pp. 1–6.

[123] J. R. Hershey, J. Le Roux, and F. Weninger, "Deep unfolding: Model-based inspiration of novel deep architectures," Nov. 2014, arXiv:1409.2574. [Online]. Available: http://arxiv.org/abs/1409.2574.

[124] A. Balatsoukas-Stimming and C. Studer, "Deep unfolding for communications systems: A survey and some new directions," in *Proc. IEEE Int. Workshop Signal Process. Syst. (SiPS)*, Oct. 2019, pp. 266–271.

[125] O. Zhao, W-S Liao, K. Li, T. Matsumura, F. Kojima, and H. Harada,

"Lazy learning-based self-interference cancellation for wireless communication systems with in-band full-duplex operations," in *Proc. IEEE 32nd Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2021, pp. 1589–1594.

[126] M. H. Attar *et* al., "Parallel APSM for fast and adaptive digital SIC in full-duplex transceivers with nonlinearity," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2022, pp. 1–5.

[127] A. Balatsoukas-Stimming, "Joint detection and self-interference cancellation in full-duplex systems using machine learning," in *Proc. 55th Asilomar Conf. Signals, Syst., Comput.*, Mar. 2021, pp. 989–992.

[128] T. Liu *et al.*, "Dynamic behavioral modeling of 3G power amplifiers using real-valued time delay neural networks," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 3, pp. 1025–1033, Mar. 2004.

[129] J. Cai, C. Yu, L. Sun, S. Chen, and J. B. King, "Dynamic behavioral modeling of RF power amplifier based on time-delay support vector regression," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 2, pp. 533–543, Feb. 2019.

[130] A. Balatsoukas-Stimming, "End-to-end learned self-interference cancellation," in *Proc. 56th Asilomar Conf. Signals, Syst., Comput.*, Mar. 2022, pp. 1334–1338.

[131] O. Ploder *et al.*, "SICNet—Low complexity sample adaptive neural network-based self-interference cancellation in LTE-A/5G mobile transceivers," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 958–972, Jun. 2022.

[132] Y. Chen *et. al.*, "MIMO full duplex radios with deep learning," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jul. 2020, pp. 1–6.

Chapter 6

[1] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

[2] F. Abbas, X. Yuan, M. S. Bute, and P. Fan, "Performance analysis using full duplex discovery mechanism in 5G-V2X communication networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 11453–11464, Aug. 2022.

[3] Z. Wang, W. Shi, W. Liu, Y. Zhao, and K. Kang, "Performance analysis of two-way full-duplex relay mixed RF/FSO system with self-interference," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 209–213, Jan. 2021.

[4] L. Shen, B. Henson, Y. Zakharov, and P. Mitchell, "Digital self-interference cancellation for underwater acoustic systems," *IEEE Trans. Circuits Syst. II: Exp. Briefs*, vol. 67, no. 1, pp. 192–196, Jan. 2020.

[5] 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.

[6] 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.

[7] D. Korpi, L. Anttila, and M. Valkama, "Nonlinear self-interference cancellation in MIMO full-duplex transceivers under crosstalk," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, pp. 1–15, Dec. 2017.

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

pp. 1–5.

[9] 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.

[10] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Advanced machine learning techniques for self-interference cancellation in full-duplex radios," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 1149–1153.

[11] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Low complexity neural network structures for self-interference cancellation in full-duplex radio," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 181–185, Jan. 2021.

[12] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Full-duplex self-interference cancellation using dual-neurons neural networks," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 557–561, Mar. 2022.

[13] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Hybrid-layers neural network architectures for modeling the self-interference in full-duplex systems," *IEEE Trans. Veh Technol.*, vol. 71, no. 6, pp. 6291–6307, Jun. 2022.

[14] M. Erdem, H. Ozkan, and O. Gurbuz, "Nonlinear digital self-interference cancellation with SVR for full duplex communication," in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC)*, Jun. 2020, pp. 1–6.

[15] M. Yilan, O. Gurbuz, and H. Ozkan, "Integrated linear and nonlinear digital cancellation for full duplex communication," *IEEE Wireless Communications*, vol. 28, no. 1, pp. 20–27, Feb. 2021.

[16] F. Jochems and A. Balatsoukas-Stimming, "Non-linear self-interference cancellation via tensor completion," in *Proc. 54th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2020, pp. 905–909.

[17] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Identification of non-linear RF systems using backpropagation," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2020, pp. 1–6.

[18] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Machine learning-based self-interference cancellation for full-duplex radio: Approaches, open challenges, and future research directions," *IEEE Open J. Veh. Technol.*, vol. 5, no. 1, pp. 21–47, Nov. 2023.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Jun. 2016, pp. 770–778.

[20] H. Zhu, Z. Cao, Y. Zhao, and D. Li, "Learning to denoise and decode: A novel residual neural network decoder for polar codes," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 8725–8738, Aug. 2020.

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

Chapter 7

[1] L. Bariah *et al.*, "A prospective look: Key enabling technologies, applications and open research topics in 6G networks," *IEEE Access*, vol. 8, pp. 174792–174820, Aug. 2020.

[2] A. M. M. Chandran and M. Zawodniok, "Measurement of internal self-

interference of USRP family of devices in full-duplex operations," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (IMTC)*, Jun. 2020, pp. 1-6.

[3] 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.

[4] 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.

[5] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Low complexity neural network structures for self-interference cancellation in full-duplex radio," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 181–185, Jan. 2021.

[6] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Full-duplex self-interference cancellation using dual-neurons neural networks," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 557–561, Mar. 2022.

[7] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Hybrid-layers neural network architectures for modeling the self-interference in full-duplex systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6291–6307, Jun. 2022.

[8] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Residual neural networks for learning the full-duplex self-interference," in *Proc. 57th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2023, pp. 1–4.

[9] M. Yilan, O. Gurbuz, and H. Ozkan, "Integrated linear and nonlinear digital cancellation for full duplex communication," *IEEE Wireless Com-*

munications, vol. 28, no. 1, pp. 20–27, Feb. 2021.

[10] F. Jochems and A. Balatsoukas-Stimming, "Non-linear self-interference cancellation via tensor completion," in *Proc. 54th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2020, pp. 905–909.

[11] M. Elsayed, A. A. A. El-Banna, O. A. Dobre, W. Shiu, and P. Wang, "Machine learning-based self-interference cancellation for full-duplex radio: Approaches, open challenges, and future research directions," *IEEE Open J. Veh. Technol.*, vol. 5, no. 1, pp. 21–47, Nov. 2023.

[12] L. Angrisani, D. Petri, and M. Yeary, "Instrumentation and measurement in communication systems," *IEEE Instrum. Meas. Mag.*, vol. 18, no. 2, pp. 4–10, Apr. 2015.

[13] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.

[14] J. Tang, C. Deng, and G. B. Huang, "Extreme learning machine for multilayer perceptron," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 4, pp. 809–821, Apr. 2016.

Chapter 8

[1] A. Balatsoukas-Stimming, "Joint detection and self-interference cancellation in full-duplex systems using machine learning," in *Proc. 55th Asilomar Conf. Signals, Syst., Comput.*, Mar. 2021, pp. 989–992.

[2] A. Balatsoukas-Stimming, "End-to-end learned self-interference cancellation," in *Proc. 56th Asilomar Conf. Signals, Syst., Comput.*, Mar. 2022, pp. 1334–1338.

[3] D. H. Kong, Y.-S. Kil, and S.-H. Kim, "Neural network aided digital self-interference cancellation for full-duplex communication over time-varying channels," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6201–6213, Jun. 2022.

[4] O. Ploder *et al.*, "SICNet—Low complexity sample adaptive neural network-based self-interference cancellation in LTE-A/5G mobile transceivers," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 958–972, Jun. 2022.

[5] Y. Chen *et. al.*, "MIMO full duplex radios with deep learning," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jul. 2020, pp. 1–6.