IEEE Open Journal of
Vehicular Technology

Received 3 October 2023; accepted 27 October 2023. Date of publication 9 November 2023;
date of current version 12 December 2023. The review of this article was coordinated by Editor Nan Cheng.

Digital Object Identifier 10.1109/OJVT.2023.3331185

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

MOHAMED ELSAYED [1,2], (Graduate Student Member, IEEE), AHMAD A. AZIZ EL-BANNA [1,3], (Member, IEEE), OCTAVIA A. DOBRE [1] (Fellow, IEEE), WAN YI SHIU [4], AND PEIWEI WANG [4]

(Invited Paper)

[1]Faculty of Engineering and Applied Science, Memorial University, St. John's, NL A1B 3X5, Canada
[2]Faculty of Engineering, Sohag University, Sohag 1646130, Egypt
[3]Faculty of Engineering at Shoubra, Benha University, Banha 13511, Egypt
[4]Huawei Technologies Canada Co., Ltd., Ottawa, ON K2K 3J1, Canada

CORRESPONDING AUTHOR: OCTAVIA A. DOBRE (e-mail: odobre@mun.ca).

This work was supported by Huawei Technologies Canada Co., Ltd., Canada.

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 article 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 article discusses the challenges of applying ML-based techniques to SIC, highlights their potential solutions, and provides a guide for future research directions.

INDEX TERMS Artificial intelligence, deep learning (DL), full-duplex (FD), machine learning (ML), neural networks (NNs), self-interference cancellation (SIC), sixth-generation (6G) wireless, support vector regressors (SVRs).

NOMENCLATURE

1T1R	One transmit and one receive.
2HLNN	Two-hidden layers neural network.
6G	Sixth-generation.
ADC	Analog-to-digital converter.
APSM	Adaptive projected subgradient method.
BPF	Band-pass filter.
BS	Batch size.
CHRNN	Channel robust neural network.
CReLU	Complex rectified linear unit.
CSID	Canonical system identification.
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.
DR	Dynamic regression.
DU	Deep unfolding.
EM	Expectation-maximization.
FD	Full-duplex.
FDD	Full 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.
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.
MWGS	Moving-window grid structure.
NN	Neural network.
NTDSVR	Nested time-delay support vector regressor.
OF-TDSVR	Output-feedback time-delay support vector regressor.
OFDM	Orthogonal frequency division multiplexing.
PA	Power amplifier.
PC	Personal computer.
PSD	Power spectral density.
QPSK	Quadrature phase-shift keying.
RBF	Radial basis function.
RF	Radio frequency.
RFFs	Random Fourier features.
RMSprop	Root mean square propagation.
RNN	Recurrent neural network.
RTDSVR	Residual time-delay support vector regressor.
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.
SoI	Signal-of-interest.
SVR	Support vector regressor.
TC	Tensor completion.
TDSVR	Time-delay support vector regressor.
TRP	Transmit and receive point.
Tx	Transmitter.
VGA	Variable gain amplifier.

I. 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], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [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 system' users [1], [2], [3], [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], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [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. Generally, the SIC can be performed in analog and/or digital domains. Analog domain cancellation can be performed passively at the radio frequency (RF), i.e., propagation level, using antenna isolation [21], beamforming [28], polarized antennas [44], circulators [45], and/or hybrid junction networks [46]. Instead, 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], [48], [49], [50], [51], [52], [53], [54], [55], [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 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], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71], [72], [73], [74], [75], [76], [77], [78], [79], [80], [81], [82], [83], [84], [85], [86], [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], [91], [92], [93], [94], [95], [96], [97], [98], [99], [100], [101], [102], [103], [104], [105], [106], [107], [108], [109], [110], [111], [112], [113], [114], [115], [116], [117], [118], [119], [120], [121], [122], [123], [124], [125], [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], [91], [92], [93], [94], [95], [96], [97], [98], [99], [100], [101], [102], [103], [104], [105], [106], [107], [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 work addresses the knowledge gap in integrating data-driven ML approaches with FD communications, applying digital SIC.

The major contributions of this work are as follows:

- We firstly introduce a general and comprehensive system model to integrate ML with FD communications.
- We have briefly reviewed the traditional approaches for SIC in FD transceivers.
- We 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.
- We have investigated the effect of using part of the output samples as features for training the SVR-based cancelers.
- We 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.
- We have devised an efficiency measure to select a suitable ML approach for SIC in FD transceivers, depending on the system requirements.
- We 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 article is organized as follows. Section II introduces the ML-based FD system model. Section III summarizes the traditional approaches for SIC in FD transceivers. Section IV reviews the up-to-date contributions that apply ML approaches for SIC. Simulation results are presented in Section V, challenges and future directions are summarized in Section VI, and finally, concluding remarks are drawn in Section VII. The detailed organization of this article is depicted in Fig. 1.

II. 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. 2. 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], [100], [101]

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

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 (1), p is an odd number, i.e., the odd-order non-linearities are only taken into account, e.g., $p \in \{3, 5\ldots, 9\}$, 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]

$$y(n) = y_{SI}(n) + y_{SoI}(n) + w(n), \quad (2)$$

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], [100], [101]

$$y_{SI}(n) = \sum_{\substack{p=1,\\ p\,odd}}^{P} \sum_{q=0}^{p} \sum_{m=0}^{M_i-1} h_{m,q,p}\, x\,(n-m)^{q}\, x^{*}\,(n-m)^{p-q}, \quad (3)$$

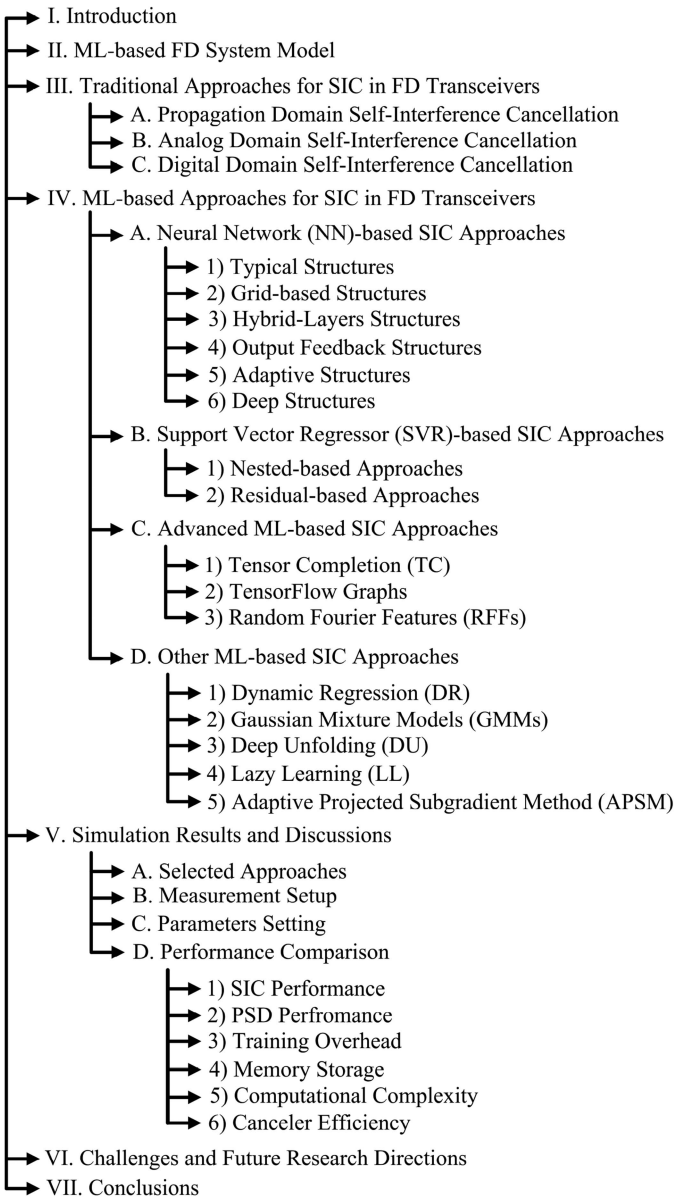


FIGURE 1. Organization of this paper.

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, we 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], [100], [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, we 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], [100], [101]:

$$\hat{\mathbf{h}} = \left(\left(\mathbf{X}^{tr}\right)^{\mathrm{H}} \mathbf{X}^{tr}\right)^{-1} \left(\mathbf{X}^{tr}\right)^{\mathrm{H}} \mathbf{y}^{tr}, \tag{4}$$

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

$$\mathbf{X}^{tr} = \begin{bmatrix} x(n) & x(n-1) & \cdots & x(n-M_i+1) \\ x(n+1) & x(n) & \cdots & x(n-M_i+2) \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ x(n+N_{tr}-M_i-1) & \cdots & \cdots & x(n+N_{tr}-2M_i) \end{bmatrix}, \tag{5}$$

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:

$$\tilde{\mathbf{y}}^{tr}_{SI,lin} = \hat{\mathbf{h}} \otimes \mathbf{x}^{tr}, \tag{6}$$

$$\tilde{\mathbf{y}}^{ts}_{SI,lin} = \hat{\mathbf{h}} \otimes \mathbf{x}^{ts}, \tag{7}$$

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 $\mathbf{x}^{tr} = [x(n) \quad x(n+1) \quad \cdots \quad x(n+N_{tr}-M_i-1)]^{\mathrm{T}}$, $\mathbf{x}^{ts} \in \mathbb{C}^{(N_{ts}-M_i)\times 1}$ 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 $\tilde{\mathbf{y}}^{tr}_{SI,lin}$ 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[1] as follows:

$$\mathring{\mathbf{y}}^{tr}_{SI,nl} = \mathbf{y}^{tr} - \tilde{\mathbf{y}}^{tr}_{SI,lin}. \tag{8}$$

Since the ML approaches are typically trained using real-valued inputs,[2] we separate the real and imaginary parts of $\mathbf{X}^{tr}$ and construct the input feature map, $\mathbf{Z}^{tr}_{nl} = [\mathbf{z}(n)\, \mathbf{z}(n+1) \cdots \mathbf{z}(n+N_{tr}-M_i-1)]^{\mathrm{T}}$, to train the non-linear canceler, with $\mathbf{z}(n) = [\Re\{x(n)\} \ldots \Re\{x(n-M_i+1)\}\ \Im\{x(n)\} \ldots \Im\{x(n-M_i+1)\}]$ 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

[1] We note 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].

[2] Without loss of generality, we assume that training the ML approaches in the system model of Fig. 2 is done using real-valued inputs; however, complex-valued inputs can also be employed, as will be discussed in the following sections.

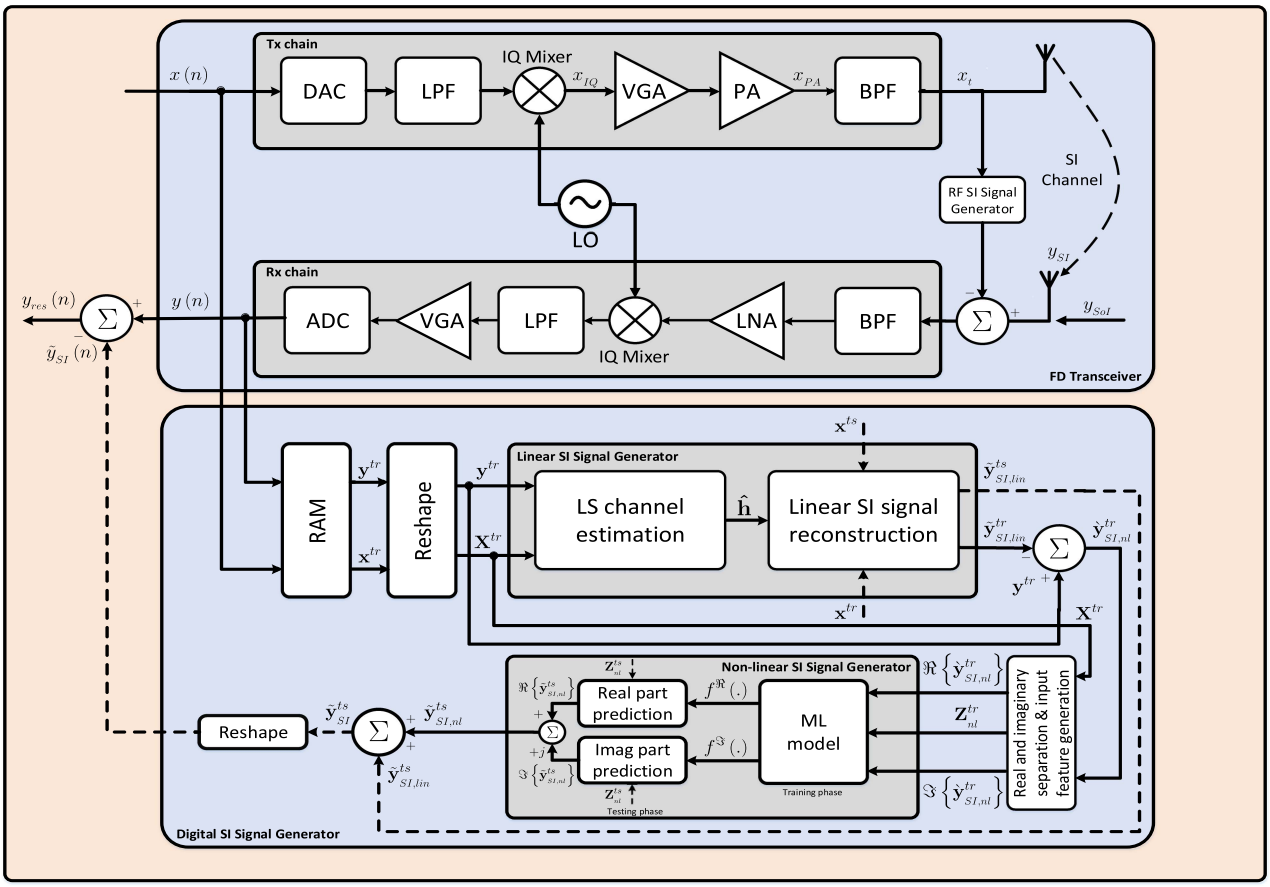


FIGURE 2. ML-based FD system model with linear and non-linear digital cancellation stages.

include a part of the output samples, as will be discussed later in Section IV. Upon constructing the input feature map, $\mathbf{Z}_{nl}^{tr}$, we separate the real and imaginary parts of $\dot{\mathbf{y}}_{SI,nl}^{tr}$ 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 utilized with $\Re\{\dot{\mathbf{y}}_{SI,nl}^{tr}\}$ and $\Im\{\dot{\mathbf{y}}_{SI,nl}^{tr}\}$ 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

$$\Re\left\{\widetilde{\mathbf{y}}_{SI,nl}^{ts}\right\} = f^{\Re}\left(\mathbf{Z}_{nl}^{ts}\right), \tag{9}$$

$$\Im\left\{\widetilde{\mathbf{y}}_{SI,nl}^{ts}\right\} = f^{\Im}\left(\mathbf{Z}_{nl}^{ts}\right), \tag{10}$$

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 description, the non-linear SI signal is obtained by summing the real and imaginary parts as

$$\widetilde{\mathbf{y}}_{SI,nl}^{ts} = \Re\left\{\widetilde{\mathbf{y}}_{SI,nl}^{ts}\right\} + j\Im\left\{\widetilde{\mathbf{y}}_{SI,nl}^{ts}\right\}. \tag{11}$$

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

$$\widetilde{\mathbf{y}}_{SI}^{ts} = \widetilde{\mathbf{y}}_{SI,lin}^{ts} + \widetilde{\mathbf{y}}_{SI,nl}^{ts}, \tag{12}$$

and the residual SI signal can be written as

$$\mathbf{y}_{res}^{ts} = \mathbf{y}^{ts} - \widetilde{\mathbf{y}}_{SI}^{ts}. \tag{13}$$

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

$$\mathcal{C}_{dB} = 10\log_{10}\left(\frac{\sum_{n=1}^{N_{ts}}|y(n)|^2}{\sum_{n=1}^{N_{ts}}|y_{res}(n)|^2}\right), \tag{14}$$

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

III. 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. 3. The following subsections briefly review such SIC approaches, discussing their advantages, disadvantages, and/or challenges.

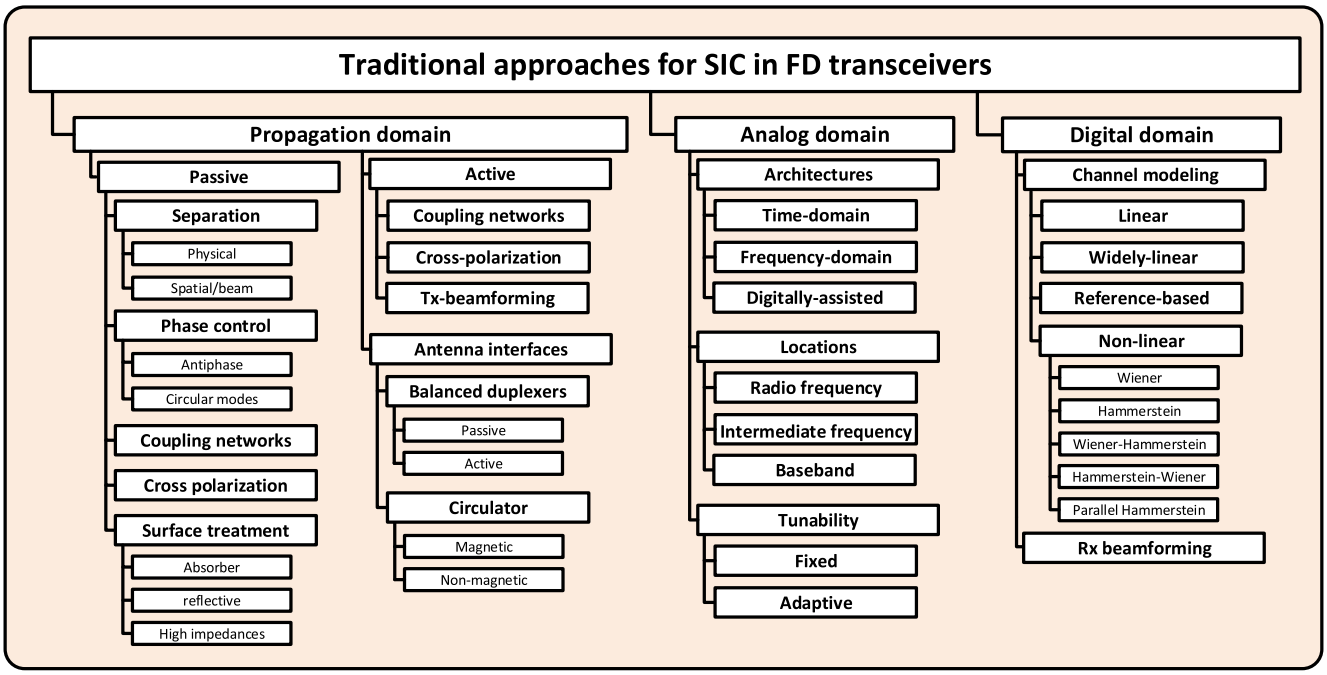


FIGURE 3. Traditional approaches for SIC in FD transceivers [43].

A. 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. 3. 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, as shown in Fig. 3. 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.

B. 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. 3 [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.

C. 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. 3. 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. 3. 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].[3]

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

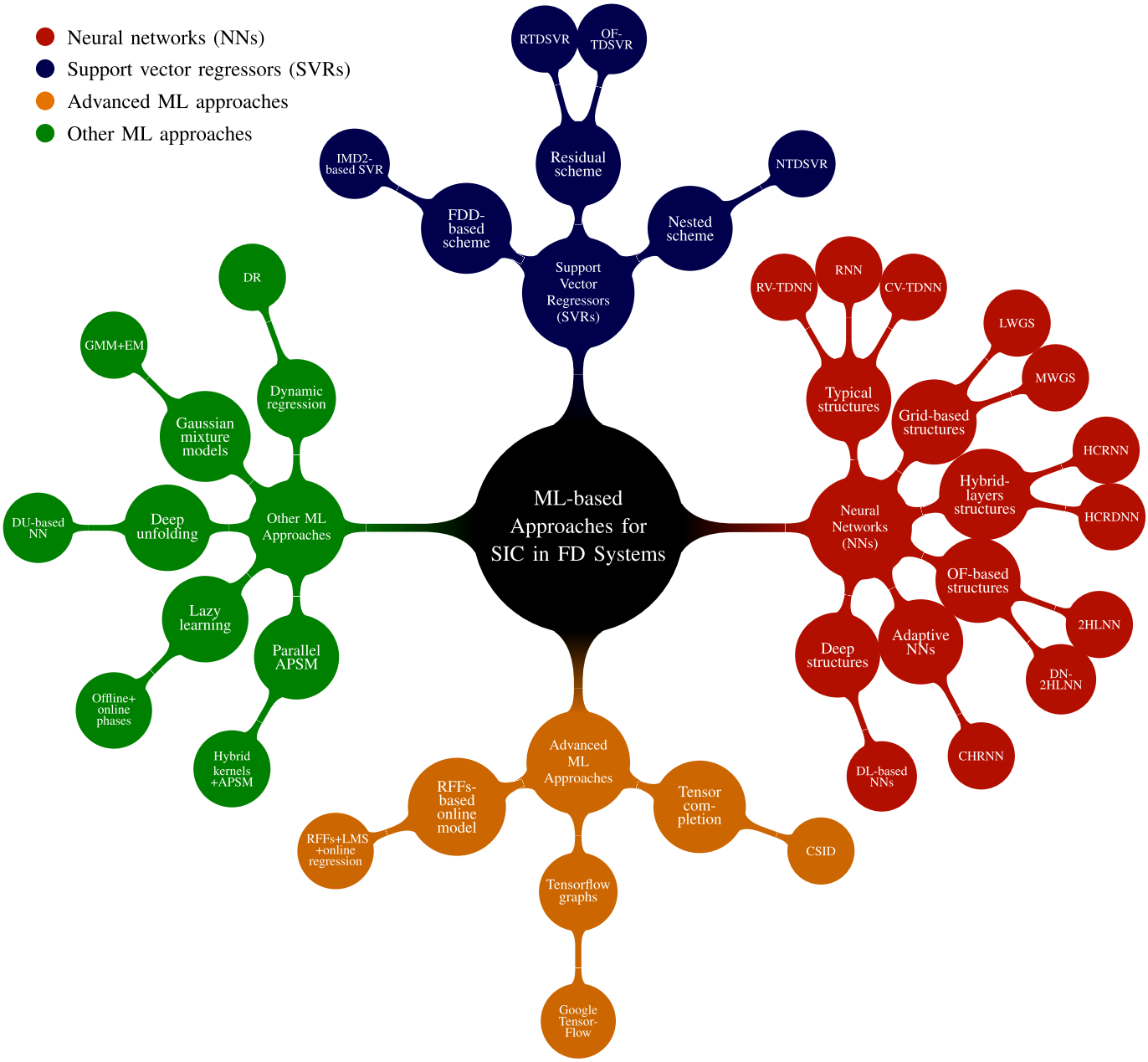


FIGURE 4. ML-based approaches for SIC in FD transceivers.

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], [96], [97], [99], [100], [101]. Given these potentials, more research efforts have recently been spurred to cancel the SI in FD transceivers using ML approaches. This article 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.

IV. ML-BASED APPROACHES FOR SIC IN FD TRANSCEIVERS

Fig. 4 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, we aim to review and summarize the up-to-date research progress in applying ML-based approaches for SIC in FD transceivers.[4]

A. 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. 4, 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.

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.[5] 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(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].[6] 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(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].

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(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].

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(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 Fig. 5(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.

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(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

[4]We 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, we 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.

[5]Throughout this article, we will use the term RV-TDNN instead of FFNN.

[6]The RV-TDNN is also investigated for SIC in FD systems in [91], [92], [93], [94].

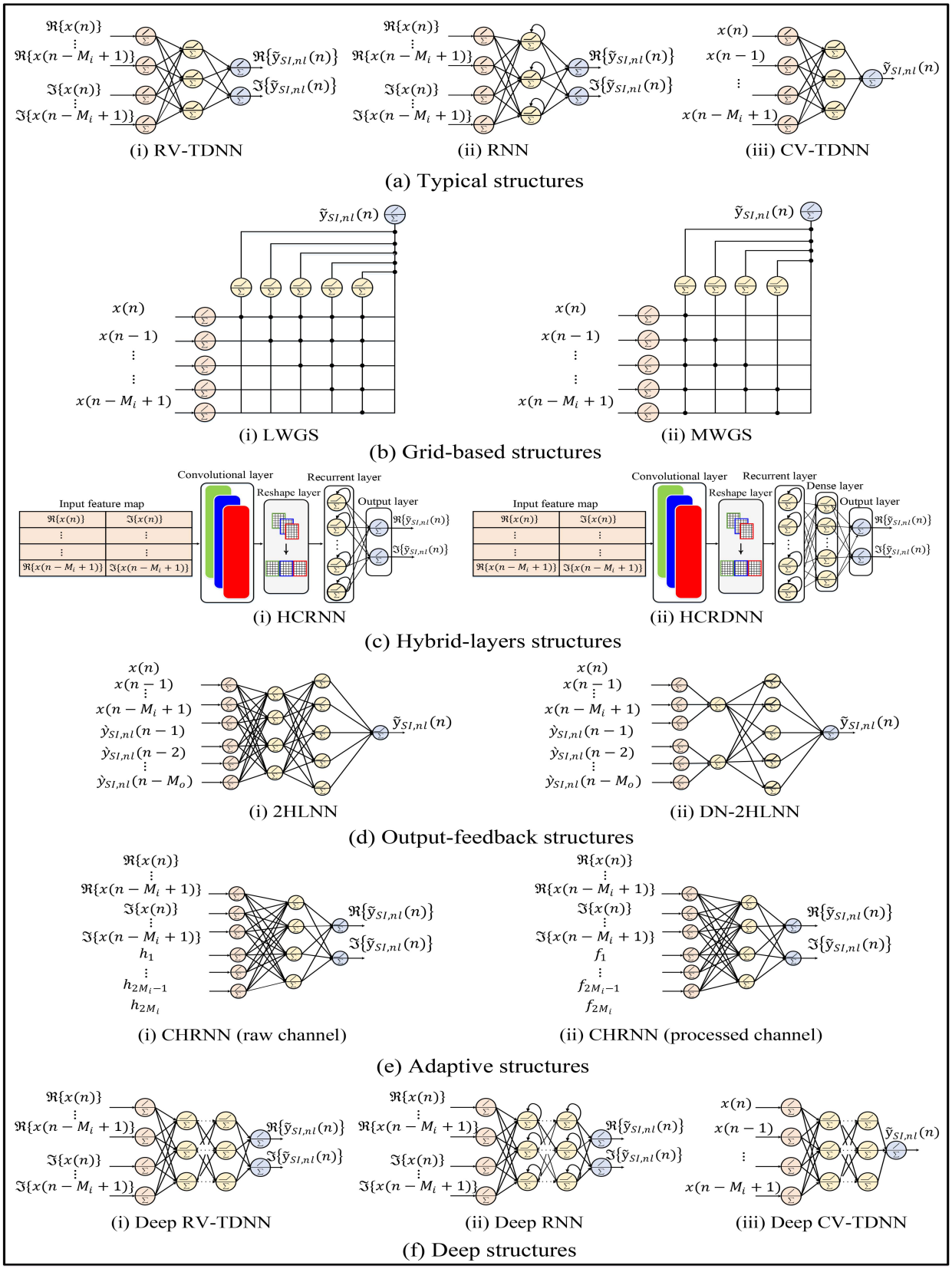


FIGURE 5. NN-based approaches for SIC in FD transceivers.

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].

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(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(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(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) 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 Fig. 5(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].

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(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], [104], [105].

B. 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.

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. 6(i), is trained using an input feature map that considers the real and imaginary parts of the current and past input samples.

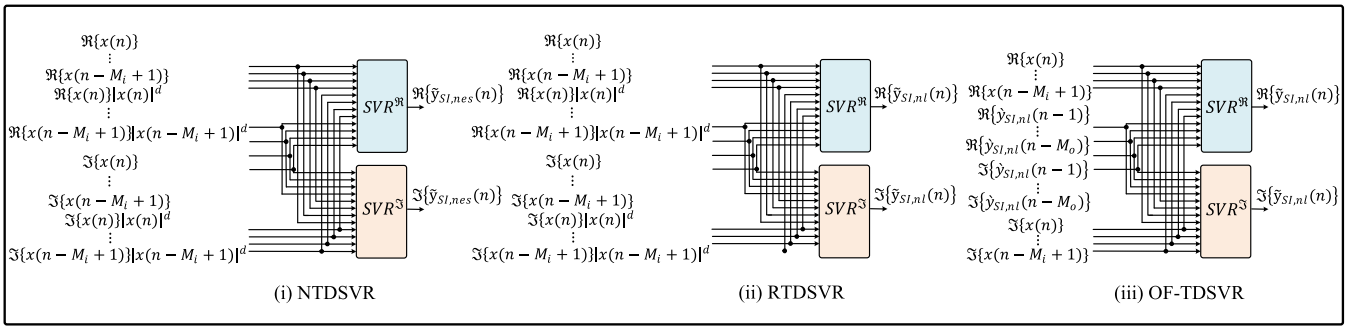


FIGURE 6. SVR-based approaches for SIC in FD transceivers. The NTDSVR is trained using $\tilde{y}_{SI,nes}$, which is generated after estimating the SI channel and performing the inverse channel filtering. However, the RTDSVR and OF-TDSVR are trained using $\tilde{y}_{SI,nl}$, which is generated after linear SI estimation and reconstruction [108].

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. 6(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].

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. 6(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].[7]

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 article*, 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. 6(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 literature benchmarks.

C. 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.

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.

[7]We note 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.

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].

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].[8]

D. 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.

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].

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.

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

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

[8] Although the RFFs are integrated with online regression in [112], they are utilized with various ML algorithms in other disciplines, such as [115], [116], [117], [118].

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) 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, we have surveyed the up-to-date contributions that apply ML-based approaches for SIC in FD transceivers, as summarized in Table 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.

V. SIMULATION RESULTS AND DISCUSSIONS

In this section, we provide a case study to compare the performance of the prominent ML approaches, surveyed in Section IV, with that of the polynomial canceler for two test setups (i.e., two training datasets) and using various dataset sizes. Specifically, we 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.

A. SELECTED APPROACHES

First, from the NN-based approaches shown on the right-hand side of Fig. 4, we 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, we 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. 4, we select the RTDSVR [108] as it is shown to outperform the NTDSVR [107], especially for low or moderate transmit power levels. Additionally, we 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 ML approaches, shown on the lower- and left-hand sides of Fig. 4, we 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, we 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.[9]

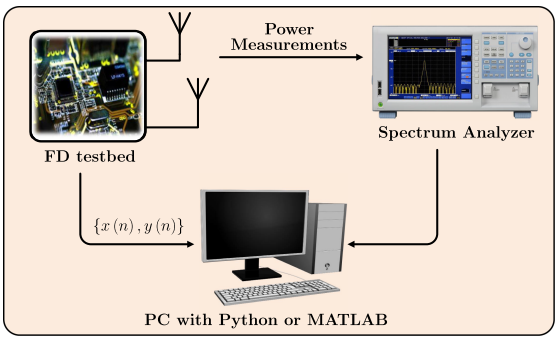


FIGURE 7. Measurement setup.

B. MEASUREMENT SETUP

The measurement setup utilized to capture the datasets employed for training the prominent ML-based approaches selected in Section V-A is described in Fig. 7. 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], [109]. The first dataset [90] applies an orthogonal frequency division multiplexing (OFDM) signal with a quadrature phase-shift keying (QPSK) modulation and 10 MHz bandwidth, while the second [96], [109] 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

[9]Up to the authors' 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 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]

TABLE 2. Measurement Setup Specifications

Unit	Parameter	First dataset [90]	Second dataset [96], [109]
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	

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.[10] Finally, for analyzing the performance of various ML-based approaches at different dataset sizes, we 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 2.

C. PARAMETERS SETTING

The goal of this work is to find the peak performance of each SI canceler, e.g., polynomial, NN, SVR, TC, and DU. In other words, we aim to find the maximum SIC that each canceler can attain. Then, we 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], we 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], we 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], we have obtained the optimum value for the memory length M_i, regularization term C, margin ϵ, along with the kernel hyperparameter, namely γ; 4) for the TC approach [109], we 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, we 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 3 and 4, respectively.

D. PERFORMANCE COMPARISON

In this subsection, we 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, we 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 V-B, and with parameter settings optimized in Section V-C.

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 Fig. 8(a) and (b), respectively.[11] From the 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.[12]

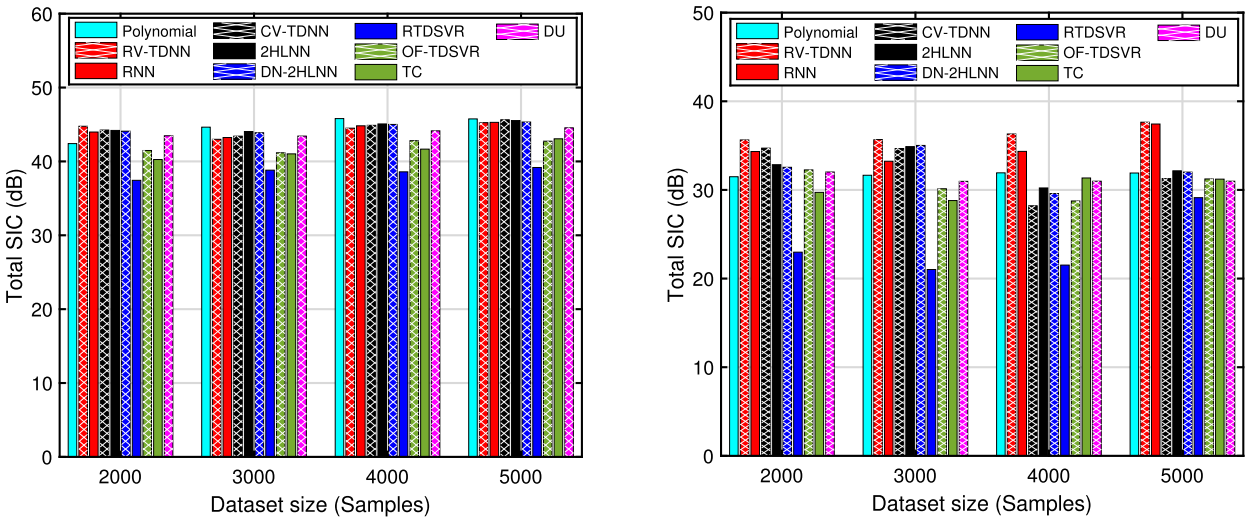
[10] We note that all ML-based SIC approaches selected for comparison in this work are implemented using Python integrated development environment, except for the TC, which is developed using the MATLAB independent development environment [109].

[11] In this work, we 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 our previous works in [99], [100], [101], we 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 work may differ from those reported in [99], [100], [101].

[12] 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, we interestingly note that the total SIC achieved in the

TABLE 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	$BS \in \{4, 6, ..., 10\}$



(a) First dataset.

(b) Second dataset.

FIGURE 8. SIC of different ML-based SI cancelers compared to the polynomial canceler over the first and second datasets.

2) PSD PERFORMANCE

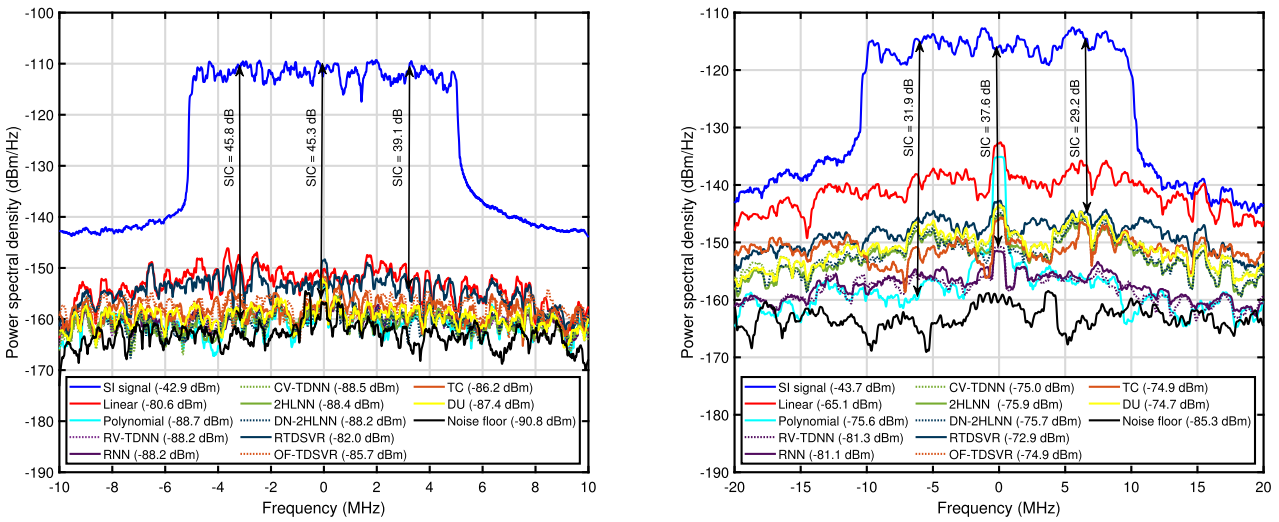
The power spectra of the residual SI signal after applying the different ML-based SIC approaches compared to that of the polynomial-based canceler when tested using the first and second datasets and with 5000 samples, as an example, is shown in Fig. 9(a) and (b), respectively. From Fig. 9(a), one can 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. 9(b) 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 comes from the fact that it can reduce the leakage of the carrier around the DC tone, as shown in Fig. 9(b) [96]. Finally, one can observe from the figures that the SIC

former is lower than that in the latter, as can be seen from the sample results in Table 5. This is due to the degradation of the linear canceler's performance with increased non-linearity.

TABLE 4. Optimal Hyperparameters for Different SIC Approaches Over the First and Second Datasets

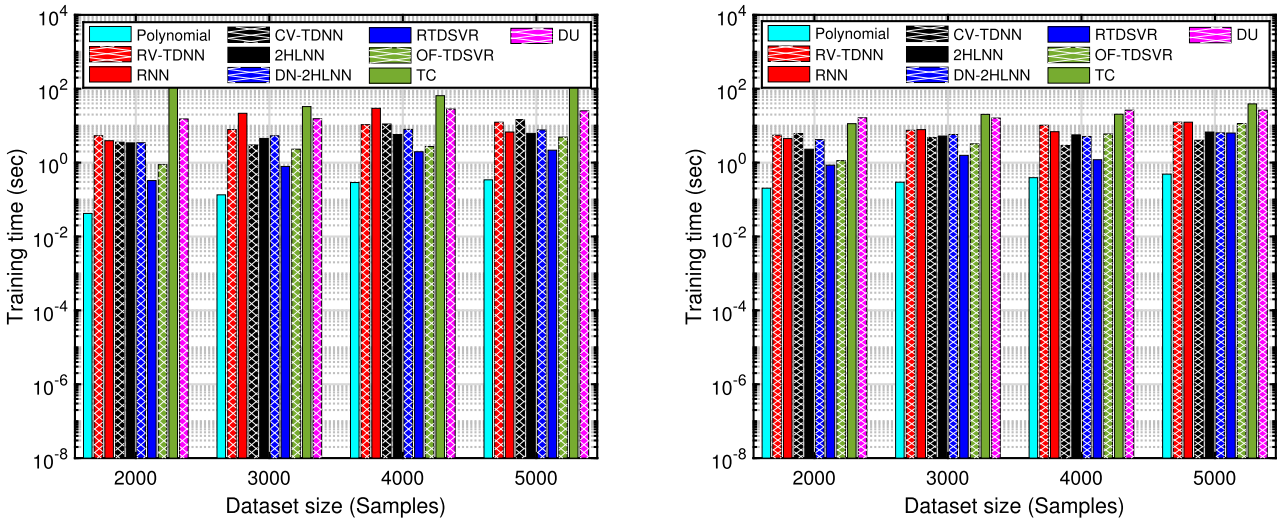
Data set	App.	Size	Method	Optimal Hyperparameters							
First dataset (implies lower transmit power, lower non-linear components, and narrower bandwidth)	Model-based	2000	Poly.	3	10	-	-	-	-	-	-
		3000		5	13	-	-	-	-	-	-
		4000		7	13	-	-	-	-	-	-
		5000		7	11	-	-	-	-	-	-
	NNs	Size	Method	M_i	M_o	n_{h1}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

Data set	App.	Size	Method	Optimal Hyperparameters							
Second dataset (implies higher transmit power, higher non-linear components, and wider bandwidth)	Model-based	2000	Poly.	9	3	-	-	-	-	-	-
		3000		9	4	-	-	-	-	-	-
		4000		9	4	-	-	-	-	-	-
		5000		9	6	-	-	-	-	-	-
	NNs	Size	Method	M_i	M_o	n_{h1}	n_{h2}	LR	BS	$Act.$	$Opt.$
		2000	RV-TDNN	3	-	96	-	0.005	22	Relu	Adam
			RNN	5	-	100	-	0.005	158	Relu	Adam
			CV-TDNN	3	-	94	-	0.005	22	CRELU	Adam
			2HLNN	7	6	2	100	0.025	158	CRELU	Adam
			DN-2HLNN	11	10	2	94	0.005	62	CRELU	Adam
		3000	RV-TDNN	3	-	93	-	0.005	22	Relu	Adam
			RNN	5	-	94	-	0.0025	62	Relu	Adam
			CV-TDNN	13	-	69	-	0.0045	62	Cardioid	Adam
			2HLNN	10	9	2	100	0.005	62	Cardioid	Adam
			DN-2HLNN	11	10	2	92	0.005	62	Cardioid	Adam
		4000	RV-TDNN	4	-	97	-	0.0025	22	Relu	Adam
			RNN	5	-	98	-	0.0025	62	Relu	Adam
			CV-TDNN	8	-	38	-	0.0045	256	Cardioid	Adam
			2HLNN	9	8	2	11	0.0045	62	Cardioid	Adam
			DN-2HLNN	6	5	2	10	0.0045	62	CRELU	Adam
		5000	RV-TDNN	6	-	100	-	0.0025	22	Relu	Adam
			RNN	6	-	90	-	0.0025	62	Relu	Adam
			CV-TDNN	12	-	89	-	0.0045	158	CRELU	Adam
			2HLNN	12	11	2	81	0.0045	62	CRELU	Adam
			DN-2HLNN	11	10	2	76	0.0045	62	CRELU	Adam
	SVRs	Size	Method	M_i	M_o	$C^{\Re}$	$\epsilon^{\Re}$	$\gamma^{\Re}$	$C^{\Re}$	$\epsilon^{\Re}$	$\gamma^{\Re}$
		2000	RTDSVR	13	-	4	0.01	0.0625	4	0.0001	0.0625
			OF-TDSVR	3	2	64	0.025	0.0625	128	0.05	0.0625
		3000	RTDSVR	12	-	4	0.01	0.0625	4	0.01	0.0625
			OF-TDSVR	5	4	32	0.0025	0.0625	64	0.05	0.0625
		4000	RTDSVR	3	-	4	0.005	0.0625	8	0.0001	0.0625
			OF-TDSVR	5	4	32	0.01	0.0625	64	0.05	0.0625
		5000	RTDSVR	12	-	4	0.005	0.0625	4	0.01	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	-	-
		2000	TC	2	-	4	32	0.0001	0.001	-	-
		3000		2	-	5	32	0.000001	0.001	-	-
		4000		2	-	5	32	0.00001	0.001	-	-
		5000		2	-	5	64	0.00001	0.001	-	-
	Other	Size	Method	P	M_i	M_o	LR	BS	$Opt.$	-	-
		2000	DU	9	12	-	0.25	4	FTRL	-	-
		3000		7	12	-	0.25	6	FTRL	-	-
		4000		7	5	-	0.25	4	FTRL	-	-
		5000		9	5	-	0.25	6	FTRL	-	-



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

FIGURE 9. PSD of different ML-based SI cancelers compared to the polynomial canceler over the first and second datasets.



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

FIGURE 10. Training time of different ML-based SI cancelers compared to the polynomial canceler over the first and second datasets.

TABLE 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

values achieved by the polynomial, RV-TDNN, and RTDSVR cancelers, as an example, match those reported in Table 5.

3) TRAINING OVERHEAD

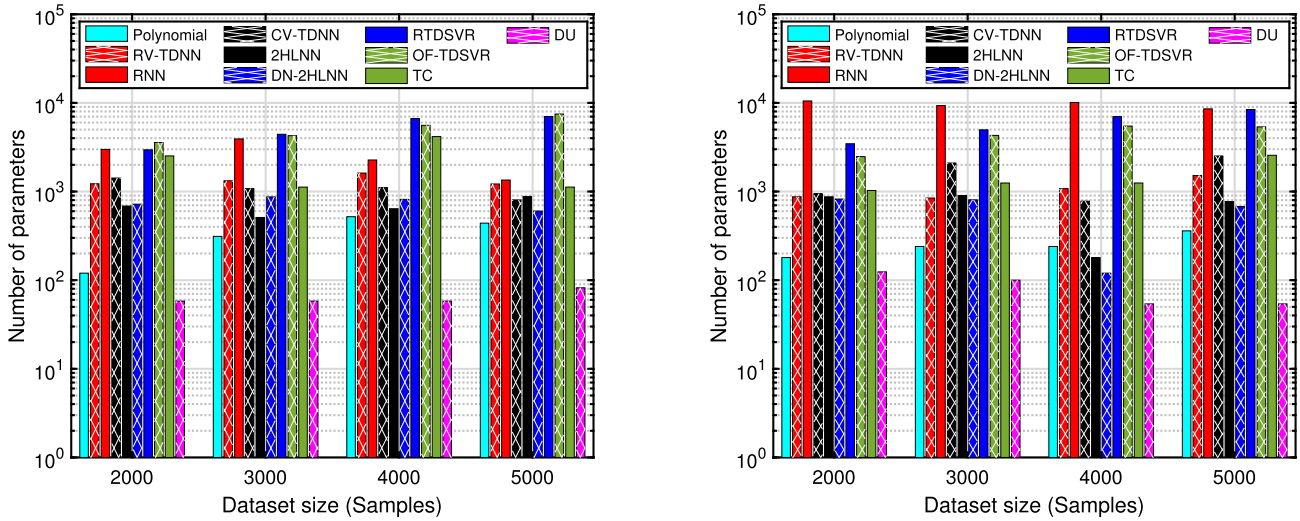
In this subsection, we assess the training time, i.e., fitting time, required by each SI canceler to complete the training process. Specifically, for the polynomial-based canceler, we evaluate the training time needed to estimate the polynomial model's coefficients based on the LS algorithm. For the NN- and DU-based cancelers, we calculate the training time as the average training time required over different random seeds. For the SVR models, we 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. 6. Finally, for the TC-based canceler, we evaluate the training time required for fitting the low-rank tensor decomposition problem. Based on the aforementioned, in Fig. 10(a) and (b), we 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.

4) MEMORY STORAGE

In this subsection, we 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 $2M_i + 2M_i\{(\frac{P+1}{2})(\frac{P+1}{2}+1)-1\}$ [90]. Further, the number of parameters of the typical RV-TDNN, RNN, and CV-TDNN is respectively evaluated as $2M_i(n_h+1)+3n_h+2$, $2M_i + n_h(n_h+5)+2$, and $2M_i + 2(M_i n_h + 2n_h + 1)$, 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 $2M_i + 2\{n_{h1}(M_i + M_o + n_{h2} + 1) + 2n_{h2} + 1\}$, and $2M_i + 2(M_i + M_o + 4n_{h2} + 3)$, with n_{h1} and n_{h2} as the number of neurons in the first and 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

TABLE 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]	$2\left\{M_i\left(\frac{P+1}{2}\right)+2\right\}$	$10M_i\left(\frac{P+1}{2}\right)+P+18$



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

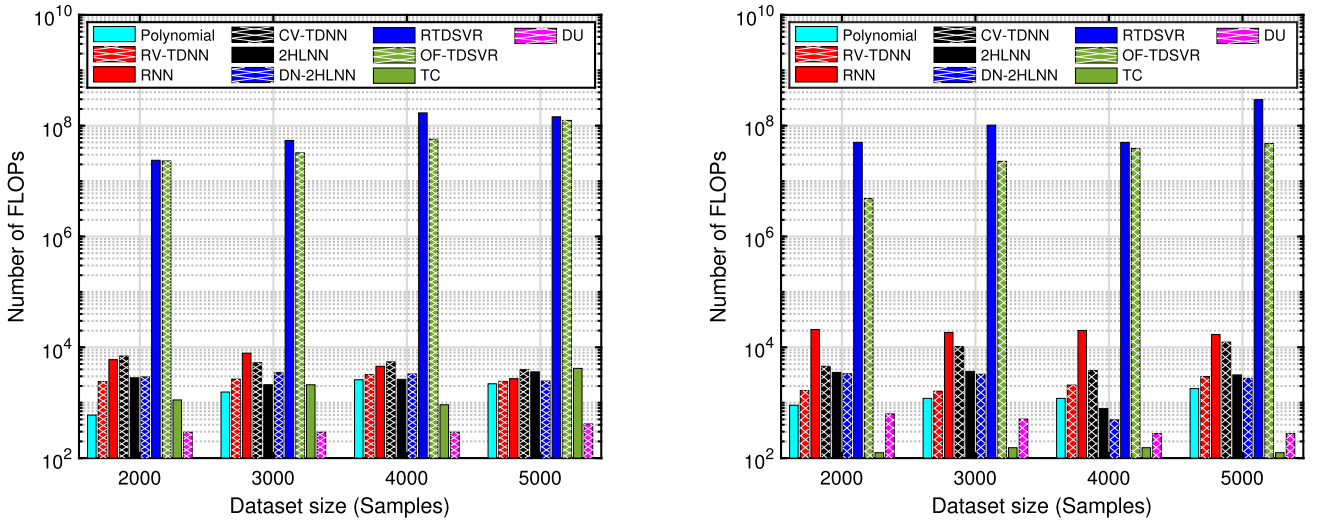
FIGURE 11. Memory storage of different ML-based SI cancelers compared to the polynomial canceler over the first and second datasets.

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 $2\{M_i(2FI+1)\}$ and $2\{M_i(\frac{P+1}{2})+2\}$, 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 6.

Based on the aforementioned, we depict the number of parameters required by the various SI cancelers when tested by the first and second datasets in Fig. 11(a) and (b), 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 notice from Fig. 11(a) and (b) 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) COMPUTATIONAL COMPLEXITY

In this subsection, we 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 $10M_i + 10M_i\{(\frac{P+1}{2})(\frac{P+1}{2}+1)-1\}-2$ [90]. Besides, the number of FLOPs of the typical RV-TDNN, RNN, and CV-TDNN are respectively evaluated as $10M_i + n_h(4M_i+5)$, $10M_i + 2n_h(n_h+\frac{9}{2})$, and $10\{M_i(n_h+1)+\frac{6}{5}n_h\}$ [90], [97]. Further, the number of FLOPs of the 2HLNN and DN-2HLNN are calculated as $10M_i + 10\{n_{h1}(M_i+M_o)+n_{h1}n_{h2}+\frac{6}{5}n_{h2}\}$ and $10M_i + 10(M_i+M_o+\frac{16}{5}n_{h2})$, respectively [101]. On the other hand, the number of FLOPs of the SVR models, i.e., RTDSVR



(a) First dataset.

(b) Second dataset.

FIGURE 12. FLOPs of different ML-based SI cancelers compared to the polynomial canceler over the first and second datasets.

and OF-TDSVR, employing an RBF kernel, are respectively evaluated in the worst case as $10M_i + 4dM_i(\frac{N_{sv}^{\Re}+N_{sv}^{\Im}}{2})Q$ and $10M_i + 4d(M_i + M_o)(\frac{N_{sv}^{\Re}+N_{sv}^{\Im}}{2})Q$, 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 $8M_i(2F + 1) - 3F - 7$ and $10\{M_i(\frac{P+1}{2}) + 2\}$ [109], [122]. A summary of the number of FLOPs utilized to asses the computational complexity of various cancelers is shown in Table 6.[13]

Based on the aforementioned, in Fig. 12(a) and (b), we 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 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].

6) CANCELER EFFICIENCY

In the previous subsections, we 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, we 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, we have devised an efficiency measure η to evaluate each canceler based on the aforementioned metrics as follows:

$$\eta = \frac{w_{\mathcal{C}}\eta_{\mathcal{C}} + w_{\tau}\eta_{\tau} + w_{\varrho}\eta_{\varrho} + w_{\mathcal{F}}\eta_{\mathcal{F}}}{w_{\mathcal{C}} + w_{\tau} + w_{\varrho} + w_{\mathcal{F}}}, \tag{15}$$

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.[14] Moreover, $\eta_{\mathcal{C}}$, η_{τ}, η_{ϱ}, and $\eta_{\mathcal{F}}$ indicate the cancellation, training, storage, and complexity efficiencies of each canceler, which can be respectively expressed as

$$\eta_{\mathcal{C}} = \frac{\mathcal{C} - \mathcal{C}_{min}}{\mathcal{C}_{max} - \mathcal{C}_{min}}, \tag{16a}$$

$$\eta_{\tau} = 1 - \frac{\tau - \tau_{min}}{\tau_{max} - \tau_{min}}, \tag{16b}$$

$$\eta_{\varrho} = 1 - \frac{\varrho - \varrho_{min}}{\varrho_{max} - \varrho_{min}}, \tag{16c}$$

$$\eta_{\mathcal{F}} = 1 - \frac{\mathcal{F} - \mathcal{F}_{min}}{\mathcal{F}_{max} - \mathcal{F}_{min}}, \tag{16d}$$

[13] In Table 6, we assume for simplicity that each RV and CV activation function costs one and two RV additions, respectively [97].

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

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, we have assessed the efficiency η for various SI cancelers over the first and second datasets in Table 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 the others for most of the test cases. One can also notice from Table 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, we 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.

VI. 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 manuscript 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.

A. 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 manuscript 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.

B. 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. 10, 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.

C. 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

TABLE 7. Efficiency η of Different ML-Based SI Cancelers Compared to the Polynomial Canceler for the First and Second datasets

Dataset	Dataset size	$w_{\mathcal{C}}$	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.		1^{st}		3^{rd}	2^{nd}					
	3000						1^{st}				2^{nd}	3^{rd}				
	4000						1^{st}				2^{nd}	3^{rd}				
	5000						1^{st}			2^{nd}	3^{rd}					
	2000	1	1	0	0	SIC and training time are the only system criteria.		1^{st}		2^{nd}	3^{rd}					
	3000						1^{st}				2^{nd}	3^{rd}				
	4000						1^{st}				2^{nd}	3^{rd}				
	5000						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						1^{st}				2^{nd}					3^{rd}
	4000						1^{st}				2^{nd}	3^{rd}				
	5000						1^{st}			2^{nd}		3^{rd}				
	2000	1	0	0	1	SIC and complexity are the only system criteria.		1^{st}		2^{nd}	3^{rd}					
	3000						1^{st}				2^{nd}	3^{rd}				
	4000						1^{st}				2^{nd}	3^{rd}				
	5000						1^{st}			2^{nd}	3^{rd}					
	2000	1	1	1	0	SIC, training time, and memory are the only system criteria.					1^{st}	3^{rd}				2^{nd}
	3000						1^{st}				2^{nd}	3^{rd}				
	4000						1^{st}				2^{nd}	3^{rd}				
	5000						1^{st}				3^{rd}	2^{nd}				
	2000	1	1	0	1	SIC, training time, and complexity are the only system criteria.		1^{st}		2^{nd}	3^{rd}					
	3000						1^{st}				2^{nd}	3^{rd}				
	4000						1^{st}				2^{nd}	3^{rd}				
	5000						1^{st}				3^{rd}	2^{nd}				
	2000	1	0	1	1	SIC, memory, and complexity are the only system criteria.					2^{nd}	3^{rd}				1^{st}
	3000						1^{st}				2^{nd}					3^{rd}
	4000						1^{st}				2^{nd}	3^{rd}				
	5000						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}				1^{st}
	3000						1^{st}				2^{nd}	3^{rd}				
	4000						1^{st}				2^{nd}	3^{rd}				
	5000						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.		1^{st}	3^{rd}	2^{nd}						
	3000							1^{st}			3^{rd}	2^{nd}				
	4000						3^{rd}	1^{st}	2^{nd}							
	5000							1^{st}	2^{nd}		3^{rd}					
	2000	1	1	0	0	SIC and training time are the only system criteria.	2^{nd}	3^{rd}						1^{st}		
	3000						1^{st}		2^{nd}		3^{rd}					
	4000						1^{st}	2^{nd}	3^{rd}							
	5000						3^{rd}	1^{st}	2^{nd}							
	2000	1	0	1	0	SIC and memory are the only system criteria.		1^{st}		2^{nd}						3^{rd}
	3000							1^{st}			3^{rd}	2^{nd}				
	4000						2^{nd}	1^{st}								3^{rd}
	5000						2^{nd}	1^{st}			3^{rd}					
	2000	1	0	0	1	SIC and complexity are the only system criteria.		1^{st}	3^{rd}	2^{nd}						
	3000							1^{st}			3^{rd}	2^{nd}				
	4000						3^{rd}	1^{st}	2^{nd}							
	5000							1^{st}	2^{nd}		3^{rd}					
	2000	1	1	1	0	SIC, training time, and memory are the only system criteria.	1^{st}	2^{nd}			3^{rd}					
	3000						1^{st}				2^{nd}	3^{rd}				
	4000						1^{st}	2^{nd}				3^{rd}				
	5000						2^{nd}	1^{st}				3^{rd}				
	2000	1	1	0	1	SIC, training time, and complexity are the only system criteria.	1^{st}	2^{nd}			3^{rd}					
	3000						1^{st}			2^{nd}	3^{rd}					
	4000						1^{st}	2^{nd}	3^{rd}							
	5000						3^{rd}	1^{st}	2^{nd}							
	2000	1	0	1	1	SIC, memory, and complexity are the only system criteria.		1^{st}		2^{nd}	3^{rd}					
	3000							1^{st}			3^{rd}	2^{nd}				
	4000						2^{nd}	1^{st}								3^{rd}
	5000						2^{nd}	1^{st}				3^{rd}				
	2000	1	1	1	1	SIC, training time, memory, and complexity are all system criteria.	1^{st}	2^{nd}				3^{rd}				
	3000						1^{st}				2^{nd}	3^{rd}				
	4000						1^{st}	2^{nd}			3^{rd}					
	5000						2^{nd}	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.

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 [132].

D. 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 VI-B. 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.

VII. CONCLUSIONS

In this paper, we have surveyed the up-to-date contributions in applying ML approaches for SIC in FD transceivers. Based on a comprehensive review, we 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, we 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, we have assessed the performance of the prominent data-driven ML-based approaches in terms of the SIC, PSD, training time, memory storage, and computational complexity and compared them with those of the model-driven approaches, e.g., polynomial-based canceler. Afterward, we 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, we 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. We 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 et al., "A prospective look: Key enabling technologies, applications and open research topics in 6G networks," *IEEE Access*, vol. 8, pp. 174792–174820, 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./Aug. 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. Calvanese 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, 2020.

[9] H. Viswanathan and P. E. Mogensen, "Communications in the 6G era," *IEEE Access*, vol. 8, pp. 57063–57074, 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, 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, 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/Jun. 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, 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," 2021, *arXiv:2101.12475*.

[20] D. C. Nguyen et al., "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 Glob. Commun. Conf.*, 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, 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, Feb. 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, T. Riihonen, V. Syrjälä, L. Anttila, M. Valkama, and R. Wichman, "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, Fourthquarter 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, Fourthquarter 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 Glob. Commun. Conf.*, 2019, pp. 1–6.

[40] H. V. Nguyen et al., "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, 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. Wireless 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 Glob. Commun. Conf.*, 2014, pp. 1001–1007.

[49] D. Korpi, T. Huusari, Y.-S. Choi, L. Anttila, S. Talwar, and M. Valkama, "Digital self-interference cancellation under nonideal RF components: Advanced algorithms and measured performance," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun.*, 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 Glob. Commun. Conf.*, 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 Glob. Conf. Signal Inform. Process.*, 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.*, 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 18th Int. Workshop Signal Process. Adv. Wireless Commun.*, 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, First Quarter 2011.

[59] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine-learning techniques in cognitive radios," *IEEE Commun. Surv. Tuts.*, vol. 15, no. 3, pp. 1136–1159, Third Quarter 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, Fourthquarter 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, Fourthquarter 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, Secondquarter 2019.
[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, Fourthquarter 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. Surv. Tuts.*, vol. 21, no. 1, pp. 393–430, Firstquarter 2019.
[72] M. Usama et al., "Unsupervised machine learning for networking: Techniques, applications and research challenges," *IEEE Access*, vol. 7, pp. 65579–65615, 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, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3039–3071, Fourthquarter 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, Fourthquarter 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, Fourthquarter 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, Jan. 2023.
[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, Dec. 2022, Art. no. 7260107.
[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 Glob. Commun. Conf.*, 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 Glob. Commun. Conf.*, 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.*, vol. 72, no. 5, pp. 6181–6193, May 2023.
[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, Jan. 2023, Art. no. 5501104.
[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 19th Int. Workshop Signal Process. Adv. Wireless Commun.*, 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. IEEE 4th Int. Conf. Commun. Inform. Syst.*, 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.*, 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. 1158–1170, 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. IEEE 52nd Asilomar Conf. Signals, Syst., Comput.*, 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. IEEE 53rd Asilomar Conf. Signals, Syst., Comput.*, 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.*, 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 Glob. Commun. Conf.*, 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.*, 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.*, 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. Netw. Conf.*, 2020, pp. 1–6.
[108] M. Yilan, O. Gurbuz, and H. Ozkan, "Integrated linear and nonlinear digital cancellation for full duplex communication," *IEEE Wireless Commun.*, 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. IEEE 54th Asilomar Conf. Signals, Syst., Comput, Comput.*, 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. IEEE Int. Conf. Comput., Netw. Commun.*, 2018, pp. 779–783.
[111] K. Wongsuphasawat et al., "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, Jul. 2022.
[113] R. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Proc. 20th Int. Conf. Adv. Neural Inf. Process. Syst.*, 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, G. Kaddoum, and G. Poitau, "Analytical guarantees for hyperparameter free RFF based deep learning in the low-data regime," *IEEE Trans. Circuits Syst. II, Express 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 Glob. Conf. Consum. Electron.*, 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*, 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, *arxXv: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.*, 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.*, 2021, pp. 1589–1594.
[126] M. H. Attar, O. Taghizadeh, K. Chang, R. Askar, M. Mehlhose, and S. Stanczak, "Parallel APSM for fast and adaptive digital SIC in full-duplex transceivers with nonlinearity," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, 2022, pp. 1–5.
[127] A. Balatsoukas-Stimming, "Joint detection and self-interference cancellation in full-duplex systems using machine learning," in *Proc. IEEE 55th Asilomar Conf. Signals, Syst., Comput.*, 2021, pp. 989–992.
[128] 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 Techn.*, 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. IEEE 56th Asilomar Conf. Signals, Syst., Comput.*, 2022, pp. 1334–1338.
[131] O. Ploder, C. Auer, C. Motz, T. Paireder, O. Lang, and M. Huemer, "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, 2022.
[132] Y. Chen, R. K. Mishra, D. Schwartz, and S. Vishwanath, "MIMO full duplex radios with deep learning," in *Proc. IEEE Int. Conf. Commun. Workshops*, 2020, pp. 1–6.

MOHAMED ELSAYED (Graduate Student Member, IEEE) received the B.Sc. degree in electronics and communications engineering from Sohag University, Sohag, Egypt, in 2014, and the M.Sc. degree in electronics and communications engineering from Assiut University, Assiut, Egypt, in 2018. He is currently working toward the Ph.D. degree in electrical engineering from Memorial University, St. John's, NL, Canada. He is also on leave from the Faculty of Engineering, Sohag University. His research interests include multiple-input multiple-output systems, wireless networks, index modulation, spatial modulation, full-duplex, and machine learning for wireless communications. He was the recipient of the Best Paper Award from the 35th National Radio Science Conference in 2018 and also the IEEE NL Graduate Scholarship in 2023.

AHMAD A. AZIZ EL-BANNA (Member, IEEE) received the Ph.D. degree in electronic and communication engineering from the Egypt-Japan University of Science and Technology, New Borg El Arab, Egypt, in 2014. He is currently an Associate Professor with Benha University, Benha, Egypt. From 2018 to 2020, he was a Postdoctoral Fellow with Shenzhen University, Shenzhen, China. From 2013 to 2014, he was a Visiting Researcher with Osaka University, Osaka, Japan. Since 2020, he has been a Postdoctoral Fellow with Memorial University, St. John's, NL, Canada. His research interests include cooperative networking, MIMO, space-time coding, IoT, underwater communication, full-duplex, and machine learning.

OCTAVIA A. DOBRE (Fellow, IEEE) received the Dipl. Ing. and Ph.D. degrees from the Polytechnic Institute of Bucharest, Bucharest, Romania, in 1991 and 2000, respectively. Between 2002 and 2005, she was with the New Jersey Institute of Technology, Newark, NJ, USA. In 2005, she joined Memorial University, Canada, where she is currently a Professor and Canada Research Chair Tier 1. She was a Visiting Professor with the Massachusetts Institute of Technology, Cambridge, MA, USA, and Université de Bretagne Occidentale, France. She has coauthored more than 450 refereed papers in her research areas, which include wireless communication and networking technologies, as well as optical and underwater communications. Dr. Dobre is the Director of Journals of the Communications Society. She was the inaugural Editor-in-Chief (EiC) of the IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY and IEEE COMMUNICATIONS LETTERS. Dr. Dobre was a Fulbright Scholar, Royal Society Scholar, and Distinguished Lecturer of the IEEE Communications Society. She was the recipient of the Best Paper Awards at various conferences, including IEEE ICC, IEEE Globecom, IEEE WCNC, and IEEE PIMRC. Dr. Dobre is an elected Member of the European Academy of Sciences and Arts, a Fellow of the Engineering Institute of Canada, and a Fellow of the Canadian Academy of Engineering.

WAN YI SHIU received the B.Eng. degree in computer engineering from McGill University, Montreal, QC, Canada, in 1994, and the M.Sc.(Eng.) degree in electrical engineering from Queen's University, Kingston, ON, Canada, in 1998. From 1997 to 2009, she was with Nortel Networks, Ottawa, ON, Canada, as a Physical Layer Digital Signal Processing software designer and FPGA firmware designer for base station cellular communication systems research and product development. Since 2010, she has been with Huawei Technologies Canada Company, Ltd., Ottawa, Canada, where she is currently a Principal Engineer. Her current research focus at Huawei is Full Duplex and machine learning Technologies. She specializes in base station cellular physical layer digital signal processing, direct digital radio frequency signal processing, software and firmware FPGA real time implementations.

PEIWEI WANG received the B.Eng. degree in computer engineering and the M.Sc. degree in computer science from Chongqing University, Chongqing, China, in 1982 and 1987, respectively. From 1995 to 1997, he was a visiting Scholar of electrical and computer science with University of Texas, Austin, TX, USA. From 1998 to 2013, he was a CDMA/DSP Designer, Architect of wireless systems with Nortel Networks and BlackBerry. In 2014, he joined Huawei Canada, where he is leading the R&D team for full duplex in wireless applications.