2024 IEEE International Conference on Communications Workshops (ICC Workshops) | 979-8-3503-0405-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCWORKSHOPS59551.2024.10615750

Hypernetwork-based Adaptive Self-Interference Cancellation for Full-Duplex Wireless Communication Systems

Sheikh Habibul Islam
Department of Electrical & Computer Engineering
University of Massachusetts Lowell,
Lowell, MA 01854
Sheikh_Islam@student.uml.edu

Xin Ma
Department of Electrical & Computer Engineering
University of Massachusetts Lowell,
Lowell, MA 01854
Xin_Ma1@uml.edu

Chunxiao Chigan
Department of Electrical & Computer Engineering
University of Massachusetts Lowell,
Lowell, MA 01854
Tricia_Chigan@uml.edu

***Abstract*— Full-duplex (FD) systems have emerged as a promising avenue to optimize temporal and spectral resource utilization by enabling simultaneous data transmission and reception on a single frequency. Nonetheless, the presence of robust self-interference (SI) signals at the receiver presents a critical challenge that necessitates effective SI cancellation strategies. Recent advancements have introduced neural networks (NN) to address this challenge, offering computational advantages over conventional polynomial models. This paper delves into the realm of leveraging Hyper Neural Networks (hyperNet) for SI cancellation (SIC), thereby exploring their unique attributes to enhance the efficiency of FD systems. The proposed method utilizes a dynamic hyperNet model to learn the non-linear characteristics of the SI channel and cancel it from the received signal. The efficacy of the proposed SIC technique is assessed using datasets derived from a MATLAB simulation platform that accurately replicates the transmission process observed in real-world scenarios. The simulation results demonstrate the superior performance of the proposed hyperNet in comparison to other state-of-art NN based SIC methods. Our proposed hyperNet-based SIC technique demonstrates its ability to autonomously adapt to more complex characteristics of the varying self-interference channel.**

Keywords*— *self-interference cancellation, time-varying channel, adaptability, hyper neural network.

I. Introduction

The rapid advancement of the Internet of Everything, which enables extensive connectivity between billions of users and devices, has prompted a significant paradigm shift towards the development of next-generation wireless networks, referred to as beyond the fifth generation (B5G) [1]. These advanced wireless networks are designed to offer exceptional reliability, minimal latency, and high data rates, reaching tens of gigabits per second.

As the demand for higher data rates in wireless communication networks continues to rise, the need for improved spectral efficiency becomes paramount. While advanced techniques like multiple-input-multiple-output (MIMO) [2] and orthogonal frequency division multiplexing (OFDM) [3] show promise in enhancing spectral efficiency, currently deployed wireless communication systems predominantly operate in half-duplex (HD) mode, resulting in compromised resource/spectrum utilization. This operational limitation leads to a capacity reduction by a factor of two, as the inherent HD constraints cannot be circumvented [4].

This project is sponsored by National Science Foundation through grant ECCS-2034530.

In the past, the practicality of in-band full-duplex (FD) wireless communication was questioned due to the significant self-interference (SI) created by a transmitter affecting its own receiver. Nevertheless, recent research [4] have revealed that it is indeed feasible to mitigate this challenge and make FD wireless systems viable. Fig.1 illustrates the classification of SI cancellation (SIC) methods within the context of FD wireless systems, as introduced in [4]. This classification is based on how the input signal is handled, distinguishing between passive, active, or a combination of both ways. As a result, active SIC, passive SIC, or a combination of both are commonly utilized to mitigate SI signals. Passive SIC is a technique employed prior to the signal's entry into the receiving antenna. It effectively mitigates SI signals by leveraging various properties associated with antennas and signal propagation. These techniques encompass the use of directional antennas, optimizing antenna spacing, and power management.

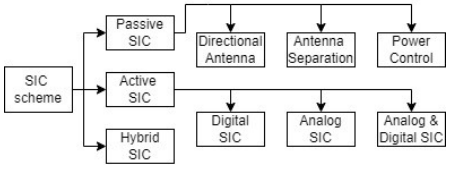

Fig. 1. Categorization of SIC approaches for FD systems [4]

The active SI mitigation approach is implemented once the signal has entered the receiving antenna. The active SIC method can be categorized into three main types: digital SIC, analog SIC, and a combination of both, as depicted in Fig. 1. Digital SIC is employed to eliminate SI arising from factors such as phase noise in the oscillator and non-linearities in the analog-to-digital converter (ADC) of the receiver. Conversely, analog SIC is utilized to address SI at the ADC stage. In many cases, a singular active SIC method, whether analog or digital, may not suffice to adequately reduce SI. Hence, a combination of both analog and digital SIC methods is often necessary to effectively mitigate SI, as noted in [5]. The choice between analog and digital SIC depends on the specific trade-offs that need to be made. Analog SIC might be preferred when simplicity, low latency, and energy efficiency are critical factors. Digital SIC is advantageous when adaptability, high cancellation accuracy, and the ability to handle complex interference scenarios are essential.

Digital SIC (DSIC) can be further classified into linear DSIC and non-linear DSIC. Linear DSIC is a technique designed to eliminate distortions in digital signals that occur

when they traverse through the environmental channel. This process aims to mitigate the impact of signal degradation caused by factors such as noise, interference, or attenuation, which can alter the shape and integrity of the signal being transmitted. On the other hand, non-linear DSIC focuses on the suppression of non-linear effects, particularly those arising from radio circuits [6]. Unlike linear distortion, which primarily affects signal amplitude and phase, non-linear distortion introduces higher-order components, such as cubic and higher-order harmonics within the transmitted signals. Non-linearities are primarily introduced by the power amplifier (PA), in-phase and quadrature-phase (IQ) imbalance of the mixer, imperfections in the transceiver's oscillators causing phase noise, and the quantization noise from the digital-to-analog converter (DAC) and ADC. The goal of non-linear digital cancellation is to reduce or eliminate these unwanted non-linear elements, ensuring that the received signals remain as faithful as possible to the originally transmitted signal.

Sophisticated memory polynomial models are a crucial requirement for effectively tackling these challenges of non-linearities, as they possess exceptional capabilities for accurately capturing the SI signal within the DSIC process. However, they suffer from high computation complexity [6]. For seeking lower complexity, a different approach has emerged as an alternative in recent years, involving the utilization of a neural network (NN) to model the non-linearities [7-11]. These NN-based DSIC techniques achieve similar SI cancellation performance with notably reduced computational complexity. Various NN-based approaches have been employed to address non-linearity in FD wireless transceiver systems. In [7], the authors utilized a basic deep NN (DNN) model for capturing the non-linearity of SI signal. In contrast, in [8], hybrid neural network architectures, specifically hybrid-layers NN, hybrid-convolutional recurrent NN, and hybrid-convolutional recurrent dense NN, were applied to capture the non-linear components of the SI signal. Additionally, they conducted performance comparisons among various recurrent neural network models. Additionally, a complex-valued feed-forward neural network (CV-FFNN) was introduced in [9] to extract non-linearities from the SI signal. Furthermore, [10] implemented a low-rank tensor completion technique known as canonical system identification (CSID) to address non-linear SI cancellation. Model based NN has been proposed in [11] where the authors used deep unfolding concept to cascade the non-linear RF system.

These Shallow NN models exhibit good performance on datasets with fixed wireless channel conditions, but their performance degrades significantly under varying channel conditions. This decline in performance is attributed to the limited training opportunities available for these simpler models. A noteworthy drawback of these existing NN-based approaches is the tendency of researchers to prioritize the perceived simplicity of their models for real-time implementation. To address this limitation, we propose a modified hyperNet architecture that employs a relaxed weight-sharing approach to extract more information from the dataset.

Contribution: This study primarily concentrates on capturing and mitigating non-linear SI signals using the noble approach, hyperNet architecture. The key strength of our hyperNet model lies in its ability to efficiently predict and adapt to evolving SI signal patterns over time, resulting in more precise and effective SI signal reduction. Unlike conventional polynomial models and other state-of-the-art NN-based methods, our proposed approach harnesses the feature of hyper networks to predict the intricate non-linear variations found in dynamic systems, thus eliminating the need for repetitive retraining cycles. The comparative analysis between a shallow model and our hyperNet is presented in Table IV, revealing favorable outcomes for our proposed model. To the best of our knowledge, our proposed model achieves a higher number of SI cancellations with similar floating-point operations per second (FLOPS) compared to other existing methods [7-11]. In order to demonstrate the high efficiency of the hyperNet, we simulated a FD system encompassing RF elements and wireless fading channel using Simulink [12]. Subsequently, we generated datasets from the simulation and employed them as inputs for the hyperNet model, thereby validating its superior performance in comparison to alternative approaches.

The subsequent sections of this paper are structured as follows. Section II, which focuses on the system model, outlines the foundational framework of FD wireless communications. In Section III elaborates the details of the hyperNet based SI canceller, presenting the innovative approach we employ for SIC. Section IV, dedicated to performance evaluations, elaborates on the methodology used in simulation, comparison of model performances. Lastly, Section V serves as the conclusion, summarizing the key findings of this article.

II. System Model of FD Wireless Communications

A FD transceiver comprises a local transmitter and a local receiver, as shown in Fig.2. The FD system is designed to employ a training-based digital cancellation technique to reduce the SI signal to the level of receiver noise. The antenna cancellation is implemented at the initial stage of the receiver system by utilizing a dual antenna setup to prevent the SI signal from saturating the analog components of the receiver, such as the Low-Noise Amplifier (LNA), Variable Gain Amplifier (VGA), and ADC. However, DSIC is used after the ADC to eliminate any remaining SI signal.

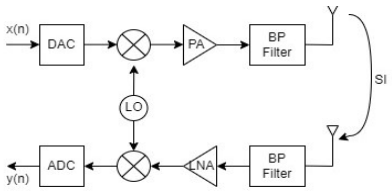

Fig. 2. Simplified illustration of a full-duplex transceiver model with a shared local oscillator (LO), omitting some components for clarity.

In Fig.2, the notation *x(n)* symbolizes the digital signal in an OFDM configuration transmitted at a discrete time point denoted as *n*. This digital signal is subject to an initial conversion into an analog format through a DAC. Following this, a process of up-conversion takes place employing an IQ mixer. In the premise where the DAC is presumed to be ideal, the digital representation of the signal at its baseband, subsequent to the incorporation of IQ imbalance attributed to the IQ mixer, can be formally articulated using the following equation [7]:

$$x_{\mathrm{IQ}}(n) = K_1 x(n) + K_2 x^*(n) \quad (1)$$

where $K_1 = \frac{1}{2}(1 + \psi e^{i\theta})$, $K_2 = \frac{1}{2}(1 - \psi e^{i\theta})$ and typically $K_1 >> K_2$. The gain and phase imbalance coefficients of the transmitter are denoted as ψ and θ respectively. Subsequently, the signal generated by the mixer is subjected to amplification through the power amplifier (PA). This process can introduce additional distortions to the transmitted signal due to inherent imperfections within the PA. The resulting signal from the PA's output can be formally delineated through the conventional parallel-Hammerstein (PH) model [7], characterized as follows:

$$x_{\mathrm{PA}}(n) = \sum_{\substack{p=1 \\ p\ odd}}^{P} \sum_{m=0}^{M} h_{\mathrm{PA},p}(m) x_{\mathrm{IQ}}(n-m) \left| x_{\mathrm{IQ}}(n-m) \right|^{p-1} \quad (2)$$

Here, $h_{\mathrm{PA},p}$ represents the PA's impulse response for the p-th order nonlinearity, and M stands for the memory length of the PA. The received x_{PA} signal makes its way to the receiver via a SI channel characterized by the impulse response $h_{\mathrm{SI}}(l)$, where $l = 0 \ldots (L-1)$ and L corresponds to overall memory of the system. The transmission from the PA inevitably permeates into the receiver through the SI signal, resulting in the emergence of the SI signal. Within the receiver of the FD node, three distinct signals become apparent: the SI signal, a noise signal, and the desired receiving signal transmitted from another FD node situated at the far-end. In this study, we adopt the assumption that neither thermal noise nor desired signals from other FD nodes are present [9]. As a consequence, the residual SI signal that remains after the RF cancellation process undergoes a sequence of operations. These operations encompass filtration through a band-pass filter (BPF), followed by amplification through a low-noise amplifier (LNA), down-conversion through an IQ mixer, and eventual digitization carried out by an ADC. This progression can be denoted mathematically as per [7], [9] :

$$y_{\mathrm{SI}}(n) = \sum_{\Delta} h_{p,q}(m) x(n-m)^q x^*(n-m)^{p-q} \quad (3)$$

where $\sum_{\Delta} = \sum_{p=odd}^{P} \sum_{q=0}^{p} \sum_{m=0}^{M+L-1}$, $h_{p,q}(m)$ is a channel impulse response having the cumulative effects of K_1, K_2, $h_{\mathrm{PA,p}}$, and h_{SI}. The objective of digital SIC is to generate an approximation of the SI signal $y_{\mathrm{SI}}(n)$, represented as $\hat{y}_{\mathrm{SI}}(n)$, wherein $\hat{y}_{\mathrm{SI}}(n)$ is a result of applying a nonlinear operation to the transmitted baseband samples. Non-linear digital canceler estimates total $h_{p,q}$ which is denoted by $\hat{h}_{p,q}$. Utilizing Eqn. (3), this canceler predicts the SI signal, $\hat{y}_{\mathrm{SI}}(n)$. Then the predicted SI signal is subtracted from the received signal. At a window length of N, the SI cancellation ratio (SICR) can be expressed as:

$$SICR_{\mathrm{dB}} = 10 \log_{10} \left(\frac{\sum_{n=0}^{N-1} |y_{SI}(n)|^2}{\sum_{n=0}^{N-1} |y_{SI}(n) - \widehat{y_{SI}}(n)|^2} \right) \quad (4)$$

III. HyperNet-based Self-Interference Canceller

Hyper Networks, also known as hyperNet is a category of NN architectures in which a smaller network is employed to create weight parameters for a larger NN referred as the main network [13]. As shown in Fig. 3, within the hyperNet system design, the more extensive network part is denoted as the main larger network. This larger network is responsible for the primary tasks. The smaller network, employed to forecast the numerical values for the weights and bias parameters of the main network, is termed the hyper smaller network.

In basic Recurrent Neural Network (RNN), the weights are tied at each time step, restricting its expressive capabilities. Dynamic hyperNet offers adaptability to generate relaxed weights for main recurrent network across the time steps. At each time step t, the smaller network receives concatenated input X_t and previous hidden state of the main network H_{t-1}. This process produces an output $\widehat{H}_t$, and this vector is employed to generate the weights for the main network at the same time step. The training of larger and smaller networks is conducted jointly through the backpropagation and gradient descent technique. Basic RNN's hidden states formula

$$H_t = \sigma(W_H H_{t-1} + W_x X_t + b) \quad (5)$$

where H_t represents the hidden state, σ denotes non-linear operation i.e. *sigmoid* or *relu*, W_H and W_x refers the weight matrices, b denotes bias, and X denotes an input sequence. These values are fixed in each time step. However, in hyperNet architecture, the weights W_h and W_x are not tied. The smaller network generates these weights for the main network so they can be different at every time step. This flexible weight-sharing approach permits us to manage the balance between model complexity and expressive power. For hyperNet architecture, H_t can be represented by [13]

$$\begin{aligned} H_t &= \sigma(W_H(\emptyset_H) H_{t-1} + W_x(\emptyset_x) + b(\emptyset_b)) \\ W_H(\emptyset_H) &= \langle W_{H\emptyset}, \emptyset_H \rangle \\ W_x(\emptyset_x) &= \langle W_{x\emptyset}, \emptyset_x \rangle \\ b(\emptyset_b) &= W_{b\emptyset} \emptyset_b + b_0 \end{aligned} \quad (6)$$

where W_H, W_x, b are computed as the inner product between the weights $W_{H\emptyset}, W_{x\emptyset}, W_{b\emptyset}$ and feature vectors $\emptyset_H, \emptyset_x, \emptyset_b$. These feature vectors are calculated by the smaller network as a function of X_t and H_{t-1}. In hyperNet architecture, the smaller network can be represented as

$$\begin{aligned} \hat{X}_t &= \begin{pmatrix} H_{t-1} \\ X_t \end{pmatrix} \\ H_t &= \sigma\left(W_{\hat{H}} \widehat{H}_{t-1} + W_{\hat{x}} \hat{X}_t + \hat{b}\right) \\ \emptyset_H &= W_{\hat{H}H} \widehat{H}_{t-1} + b_{\hat{H}H} \\ \emptyset_x &= W_{\hat{H}x} \widehat{H}_{t-1} + b_{\hat{H}x} \\ \emptyset_b &= W_{\hat{H}b} \widehat{H}_{t-1} \end{aligned} \quad (7)$$

where $W_{\hat{H}}, W_{\hat{x}}, \hat{b}$ are denoted as weights and bias of the smaller network. The feature vectors $\emptyset_H, \emptyset_x, \emptyset_b$ are of dimension of $N_{\emptyset}$, which is smaller than the hidden state size of the main network (N_H) and the smaller network ($\widehat{N}_H$). Typically, $\widehat{N}_H < N_H$. Output of the main network is projected on the feature vectors by the linear network, which later help to generate the weights for the main larger network.

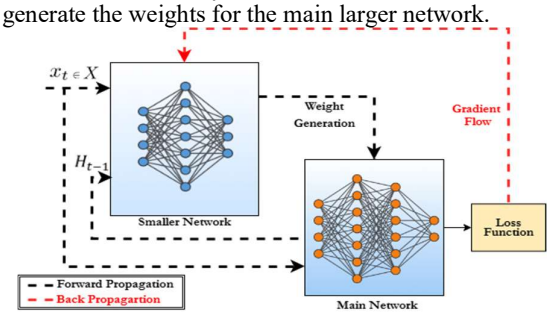

Fig. 3. Simplified Version of HyperNet Architecture

However, the exact implementation of this hyperNet architecture will cause memory overflow problem. To address the issue, modifications are necessary so that the hyperNet architecture can become more memory efficient and scalable. In the modified design, a weight scaling vector, $d(\emptyset)$ is

introduced. This vector is obtained through a linear projection of $\emptyset$. The modified $W(\emptyset)$ can be represented as

$$W(\emptyset) = W\big(d(\emptyset)\big) = \begin{pmatrix} d_0(\emptyset)W_0 \\ d_1(\emptyset)W_1 \\ \dots \\ d_{N_H}(\emptyset)W_{N_H} \end{pmatrix} \quad (8)$$

Transforming $W(\emptyset)$ into $W(d(\emptyset))$ proves to be a beneficial tradeoff, leading to a reduction in the memory demands of the hyperNet architecture. The row level operation in Eqn. (8) can be replaced with element-wise multiplication, resulting in increased practical efficiency. The modified memory efficient version of Eqn.(5) can be refered as

$$H_t = \sigma(d_H(\emptyset_H) \odot W_H H_{t-1} + d_x(\emptyset_x) \odot W_x X_t + b(\emptyset_b)) \quad (9)$$

where $d(\emptyset) = W_\emptyset \emptyset$, $b(\emptyset) = W_{b\emptyset}\emptyset_b + b_0$ and $\odot$ denotes element-wise multiplication.

A. HyperNet Non-Linear Canceler

The SI signal can be divided into two components: a linear part and a non-linear part. This can be represented as:

$$y_{\mathrm{SI}}(n) = y_{\mathrm{SI_{lin}}}(n) + y_{\mathrm{SI_{nl}}}(n) \quad (10)$$

In Eqn. (10), the component $y_{\mathrm{SI_{lin}}}(n)$represents the linear part of the SI signal. This pertains to the specific term in the sum of Eqn. (3) where *p = 1* and *q = 1*. On the other hand, the component $y_{\mathrm{SI_{nl}}}(n)$ encompasses all the remaining terms that contribute to non-linearity. We employ conventional linear cancellation techniques to create an approximation of $y_{\mathrm{SI_{lin}}}(n)$, which we will symbolize as $\hat{y}_{\mathrm{SI_{lin}}}(n)$. This approach treats the $y_{\mathrm{SI_{nl}}}(n)$ signal, which is considerably weaker, as unwanted interference, resembling noise. Subsequently, we intend to restore $y_{\mathrm{SI_{nl}}}(n)$ through hyperNet. To detail the linear cancellation process: the linear canceller initially calculates channel impulse response $\hat{h}_{1,1}$ using established methods of least-squares channel estimation [6], Subsequently, this $\hat{h}_{1,1}$ value is employed to formulate the estimation $\hat{y}_{\mathrm{SI_{lin}}}(n)$ in the subsequent manner [6]:

$$\hat{y}_{\mathrm{SI_{lin}}}(n) = \sum_{m=0}^{M+L-1} \hat{h}_{1,1}(m)x(n-m) \quad (11)$$

Once the estimation of the linear component of the SI signal is obtained, it is subsequently subtracted out from the primary SI signal. This subtraction operation results in a residual signal, which comprises the non-linear portions of the SI signal. Our hyperNet's objective is to reconstruct individual samples of $y_{\mathrm{SI_{nl}}}(n)$ based on the relevant subset of *x* that these particular $y_{\mathrm{SI_{nl}}}(n)$ samples are dependent upon, as indicated in Eqn. (10). We employ a hyperNet architecture shown in Fig. 3. This NN comprises a single layer of nodes that transmit information in one direction. Specifically, it has a total of 2 times the sum of *L* and *M* input nodes. These input nodes correspond to the real and imaginary parts of the *(M + L)* distinct time-shifted versions of *x*, as expressed in Eqn. (3). Additionally, the neural network contains two output nodes, which represent the real and imaginary parts of the desired $y_{\mathrm{SI_{nl}}}(n)$ samples.

IV. Performance Evaluations

In this section, we elaborate the simulations conducted for the FD wireless system, the configuration of the hyperNet model, and the evaluation of the proposed hyperNet-based cancellers. This evaluation is carried out in the context of suppressing the SI signal within a simulated FD system.

A. Simulation for FD System

In WLAN applications, creating an effective Simulink model requires the incorporation of three key components: an OFDM signal generator, transceiver RF elements, and an accurate wireless channel representation. Creating OFDM baseband signals need random bits generator, type of modulation, and IFFT in Simulink. This baseband signal will pass through simulated RF elements employed by the RF Blockset in Simulink [12] [14]. These toolboxes in Simulink facilitates the creation and oversampling of a baseband wireless signal. The RF elements comprises of several components such as IQ mixer, Power Amplifier (PA), Low Noise Amplifier (LNA), and Variable Gain Amplifier (VGA) essentially. These elements are characterized by gain, second-order intercept point IP2, third-order intercept point IP3, noise level, and order. We have set parameters for IQ mixer (*IP2 = 47, IP3 = inf, order = even and odd*), PA (*IP3 = 47, order = odd*), LNA (*IP2 = 7, IP3 = 10, order = even and odd*), VGA (*IP3 = 47, order = odd*). Finally, Simulink offers precise simulation tools for various wireless fading channels, allowing for the adjustment of parameters like path delays, gains, Doppler shifts, and phases to simulate diverse scenarios in wireless communication outlined in Tables I and II. Due to the page limit, we will present the details of our Simulink Simulation platform in our future publication.

TABLE I
Parameters for Rayleigh Fading Channel Modeling

Config	Delay	Gain	Doppler	AWGN
1	$[0\ 0.4\ 0.6]*10^{-1}$	[0 -30 -33]	0.001	25
2	$[0\ 0.4\ 0.6]*10^{-2}$	[0 -30 -33]	0.001	25
3	$[0\ 0.6\ 0.8]*10^{-2}$	[0 -30 -33]	0.001	25
4	$[0\ 0.4\ 0.6]*10^{-4}$	[0 -30 -33]	0.001	25
5	$[0\ 0.4\ 0.6]*10^{-6}$	[0 -30 -45]	10	25

TABLE II
Parameters for Rician Fading Channel Modeling

Config	Delay	Gain	Doppler	AWGN
1	$[0\ 0.4\ 0.6]*10^{-8}$	[0 -45 -48]	0.001	25
2	$[0\ 0.4\ 0.6]*10^{-9}$	[0.7 -45 -48]	0.001	25
3	$[0\ 0.4\ 0.6]*10^{-9}$	[2 -45 -48]	0.001	25
4	$[0\ 0.4\ 0.6]*10^{-9}$	[5 -45 -48]	0.01	25
5	$[0\ 0.4\ 0.6]*10^{-9}$	[10 -45 -48]	0.01	25

B. HyperNet Model Configuration

In this section, we meticulously present both the commonalities and disparities between our simulated datasets and provided datasets [6]. Furthermore, we emphasize the rationale behind the development of new datasets to illustrate the originality of our research.

1) Training Datasets from Real TestBed: According to the author of [6], their dataset consists of QPSK-modulated OFDM signals which has following attributes as Table I. Their transmitted OFDM frame consists of approximately 20,000 baseband samples, with 90% used for training and the remaining 10% used for testing. They employed a two-antenna setup that provides a passive analog suppression of 53 dB. Active analog cancellation is not used since their achieved passive suppression is sufficient for the work of digital SIC.

2) Training Datasets from Simulations: We conducted simulations using MATLAB Simulink to replicate a similar system architecture. This simulation generates a signal employing QPSK modulation with a 10 MHz frequency and an average transmit power of 10 dBm. The signal uses OFDM with 1024 sub-carriers and is sampled at a rate of 20 MHz. The resulting dataset consists of approximately 30,000 time-domain baseband samples. We divided this dataset into two distinct parts for different purposes. It is important to note that the simulation used in this work provided a significant 53 dB passive analog RF cancellation similar as in [6]. This degree of cancellation was attained using the component known as "Antenna" within the RF Blockset. Consequently, in our simulation, we did not employ any additional components as active analog cancellation techniques similar as in [6]. The passive analog suppression, combined with proposed digital cancellation methods, was sufficient to reduce the SI signal's power to the level of the receiver's background noise.

3) Comparison of Both Datasets: The dataset provided by the author in reference [6] presents a fixed channel condition, but they did not provide explicit details regarding their wireless channel configuration in their research. One major distinction between their study and ours lies in our ability to adjust the wireless channel parameters to closely align with real-world scenarios. This enables us to train and evaluate our model across various scenarios. Additionally, we can train our model under specific wireless channel conditions and assess its performance in the presence of changing channel conditions. Our generated dataset offers greater versatility from research aspect. Table I and II represent the datasets which have varying wireless fading channels configuration.

There are some other parameters which reflect the real-life scenarios such as rural area or urban area. In a fading channel model that reflects rural area behavior, one parameter that is commonly adjusted to capture the characteristics of rural environments is the Rician K-factor. In a rural area, especially in scenarios where there is a clear line of sight between the transmitter and receiver, the dominant line-of-sight component can be relatively strong compared to the scattered multipath components. This is because rural areas often have fewer obstacles and buildings that can cause significant multipath propagation.

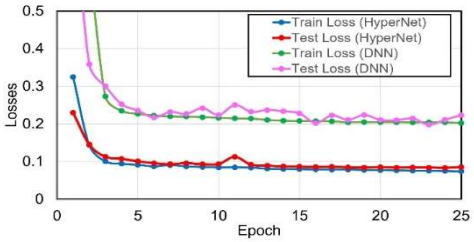

Fig. 4. MSE in the training and testing phase for the proposed hyperNet and other NN architecture.

C. Comparison of model performance

In this subsection, the evaluation of proposed hyperNet-based SI cancellers and the state-of-the-art counterparts takes place within the context of suppressing SI signal in a simulated FD system. The performance evaluation encompasses several key aspects, including the model training performance, the SI reduction performance of the modeled SI signal, and the performance under the varying wireless fading channel conditions. The hyperNet model was implemented using the PyTorch framework. The underlying optimization algorithm for training was Adam, and the chosen loss function was mean squared error. A learning rate of $\lambda = 0.004$ was utilized, with mini-batches containing 32 samples each. To ensure a fair comparison with the other NN models, our hyperNet was designed with an input layer of $2(M + L) = 26$ units. Following table III, the main network's hidden layer size, $N_H = 8$, and the hyper smaller network's hidden layer size, $\widehat{N}_H = 5$ are set. According to [13], our system's computational performance, measured in FLOPS, is estimated at approximately 3500. Notably, FLOPS serves as a widely accepted standard for evaluating performance in the field of SI cancellation, as established by previous works [6-11]. Significantly, our study demonstrates that not only do our FLOPS match those of other studies, but our cancellation limit excels, outperforming papers with similar FLOPS values.

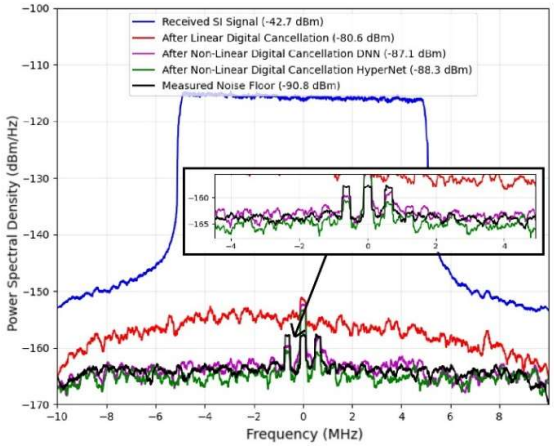

Fig. 5. Power spectral densities of the SI signal, the SI signal after linear cancellation, as well as the SI signal after non-linear cancellation using both the deep neural network and hyper network. We also show the measured noise floor for reference.

TABLE III
Hyperparameter Optimization

Input Size	N_H	$\widehat{N}_H$	$N_\emptyset$	FLOPS
26	8	5	2	~3500
26	8	6	2	~3800
26	10	6	2	~4500
26	10	8	2	~5500

1) Training Performance: Fig.4 illustrates the Mean Squared Error (MSE) performance during both the training and testing phases of the proposed hyperNet architecture, in comparison to a Deep Neural Network (DNN) architecture [6]. The MSE serves as an error metric for evaluating the difference between the predicted SI signal and the ground-truth. It is evident from Fig.4 that the MSE values differ 10% between the architectures in both the training and testing phases. Moreover, it is apparent that both the suggested hyperNet and DNN do not display indications of overfitting. They consistently perform well during both the training and testing stages. Notably, our hyperNet model excels at capturing more intricate behaviors within the FD system. Furthermore, the graph indicates that the proposed hyperNet consistently converge faster and reaches lower errors in both training and testing. This convergence underscores the

effectiveness of the solution offered by the proposed hyperNet architectures.

2) SI Reduction Performance: In this subsection, we analyze the Power Spectral Density (PSD). As observed in Fig.5, the linear SI canceller effectively reduces the SI signal power by approximately 38 dB, reducing it from -42.7 dBm to -80.60 dBm. Furthermore, the DNN canceller achieves an additional 6.6 dB of cancellation, lowering the power of the residual SI signal from -80.60 dBm to -87.1 dBm, bringing it very close to the receiver's noise floor (approximately 3 dB above the receiver's noise floor). A similar outcome is achieved by the proposed hyperNet canceller, which cancels the SI signal after the linear canceller by 7.7 dB, making it closely resemble the receiver's background noise level, as depicted in the inset graph of Fig.5.

3) Performance on Varying Wireless Channels: In this subsection, we undertook an evaluation process involving several trained models. To test and evaluate the trained models, we use new datasets where the fading channel parameters were slightly varied, as outlined in Tables I, and II. To ensure a fair and consistent comparison between our proposed hyperNetand DNN model [6], we ensured that both models were trainedusing the same dataset specified in Table IV. For further clarity, in Table IV, the column labeled "config" denotes thespecific parameter conditions that align with those detailed inTables I and II. This approach was implemented to maintain consistency and fairness in our comparative analysis of the models.

TABLE IV
Performance Comparison

Fading Channel	Config	Trained and Tested / Tested with Pretrained Model	Hyper Net (dB)	DNN (dB)
Rayleigh	1	Tested with Pretrained Model	6.8	6.2
	2	Tested with Pretrained Model	7.1	6.8
	3	Tested with Pretrained Model	6.3	6.1
	4	Trained and Tested	5.4	5.1
	5	Tested with Pretrained Model	1.5	1.2
Rician	1	Trained and Tested	7.7	2.9
	2	Tested with Pretrained Model	7.6	2.85
	3	Tested with Pretrained Model	6.7	2.6
	4	Tested with Pretrained Model	6.3	2.5
	5	Tested with Pretrained Model	5.8	2.35

Based on [15], the characteristics of fading channels are influenced by variations in channel gain and delay values. Fading channels exhibit a greater degree of variability in real-time conditions, rendering them highly susceptible to alterations in the environment's specific parameters. Notably, varying the distance between the transmitter and receiver may likely change the properties of the wireless fading channel. Our model outperforms the conventional DNN in Rayleigh fading channel and Rician fading channel, as indicated in table IV [6]. Both models undergo training and testing under the Rayleigh Fading channel configuration 4. Following this, the model is saved and applied directly to the remaining Rayleigh fading datasets. It becomes evident that across all aspects of dataset testing, our hyperNet consistently outperforms the DNN model. This performance advantage is even more pronounced in the context of the Rician Fading channel, where utilizing the pretrained model leads to significant enhancements in the detection of non-linear characteristics in the SI signal.

V. Conclusion

In this research paper, we introduce a novel approach for enhancing signal quality for FD wireless systems in the presence of self-interference. Our method combines a hyperNet with adaptive SI cancellation to effectively mitigate interference effects in dynamic communication channels. Our model excels at accurately predicting the complex interference factor. The proposed SIC technique significantly reduces interference to the point of being indistinguishable from background noise. Importantly, our approach offers a reasonable computational complexity compared to a polynomial based SIC scheme and other NN models with adaptive extensions [6 - 9]. To assess the effectiveness of our digital interference cancellation scheme in real-world scenarios, we conducted simulations in a dynamic wireless fading channel environment. Our findings demonstrate that our proposed scheme can remarkably diminish interference power to the level of background noise in situations where interference sources change over time. Furthermore, our results underscore the importance of appropriately processing estimated interference channel characteristics, particularly in changing environments with multiple signal paths causing interference.

References

[1] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6g networks: Use cases and technologies," IEEE Communications Magazine, vol. 58, no. 3, pp. 55–61, 2020.

[2] Z. Zhang, X. Wang, K. Long, A. V. Vasilakos, and L. Hanzo, "Largescale mimo-based wireless backhaul in 5g networks," IEEE Wireless Communications, vol. 22, no. 5, pp. 58–66, 2015.

[3] R. v. Nee and R. Prasad, OFDM for wireless multimedia communications. Artech House, Inc., 2000.

[4] M. Amjad, F. Akhtar, M. H. Rehmani, M. Reisslein, and T. Umer, "Fullduplex communication in cognitive radio networks: A survey," IEEE Communications Surveys & Tutorials, vol. 19, no. 4, pp. 2158–2191, 2017.

[5] M. Duarte, A. Sabharwal, V. Aggarwal, R. Jana, K. K. Ramakrishnan, C. W. Rice, and N. Shankaranarayanan, "Design and characterization of a full-duplex multiantenna system for wifi networks," IEEE Transactions on Vehicular Technology, vol. 63, no. 3, pp. 1160–1177, 2013.

[6] A. Balatsoukas-Stimming, "Non-linear digital self-interference cancellation for in-band full-duplex radios using neural networks," in 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2018, pp. 1–5.

[7] ——, "Joint detection and self-interference cancellation in full-duplex systems using machine learning," in 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021, pp. 989–992.

[8] 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 Communications Letters, vol. 26, no. 3, pp. 557–561, 2021.

[9] ——, "Low complexity neural network structures for self-interference cancellation in full-duplex radio," IEEE Communications Letters, vol. 25, no. 1, pp. 181–185, 2020.

[10] F. Jochems and A. Balatsoukas-Stimming, "Non-linear self-interference cancellation via tensor completion," in 2020 54th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2020, pp. 905–909.

[11] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Identification of non-linear rf systems using backpropagation," in 2020 IEEE International Conference on Communications Workshops (ICC Workshops). IEEE, 2020, pp. 1–6.

[12] The MathWorks, Inc. (2022). MATLAB version: 23.2.0.2489961(R20-23b). Accessed: January 01, 2023.

[13] D. Ha, A. Dai, and Q. V. Le, "Hypernetworks," 2016.

[14] The MathWorks, Inc. (2022). RF Blockset version: 23.2 (R2023b). Accessed: January 01, 2023.

[15] D. Chizhik, J. Du, and R. A. Valenzuela, "Universal path gain laws for common wireless communication environments," IEEE Transactions on Antennas and Propagation, vol. 70, no. 4, pp. 2928–2941, 2021