Cascaded Models for Cancellation of Air-Induced Passive Intermodulation in MIMO Transceivers

Vesa Lampu*, Lauri Anttila*, Matias Turunen*, Marko Fleischer†, Jan Hellmann†, and Mikko Valkama*
*Department of Electrical Engineering, Tampere University, 33720 Tampere, Finland
†Nokia Mobile Networks, D-89081 Ulm, Germany
E-mail: vesa.lampu@tuni.fi

***Abstract*—This paper describes and proposes cascaded models for the identification and cancellation of air-induced passive intermodulation (PIM) in frequency division duplex (FDD) multiple-input multiple-output (MIMO) systems. The air-induced PIM is interference which stems from nearby metallic objects outside the transceiver system. Previous works on the topic have been targeting a limited subset of the possible scenarios present in such systems, namely 2x2 MIMO operation with two parallel streams and two component carriers (CCs), totaling four distinct signals. We demonstrate that the recently proposed cascaded model utilizing spline-interpolated lookup tables for nonlinearity modeling is capable of handling the aforementioned case with low computational cost, even when channels with memory effects are considered. Furthermore, we expand the 2x2 case to include more transmit streams and CCs and develop a multi-input Hammerstein-type model with polynomial basis functions and a global filter, and accompany it with a low-complexity gradient descent-based parameter adaptation to cancel PIM in such more evolved scenarios. We evaluate the performance of the developed canceller numerically, where favorable cancellation results are shown.**

***Index Terms*—Passive intermodulation, rusty-bolt, nonlinear distortion, digital cancellation**

I. Introduction

Since the 4G Advanced/LTE, the 3GPP specifications have included carrier aggregation (CA) technology, where ever wider bandwidths are made available by dynamic allocation of the spectral resources [1]. Crucially, this spectral allocation of the component carriers (CCs) can be done contiguously or non-contiguously. In the latter case, if frequency division duplex (FDD) system is employed, and the duplexing distance is narrow, such as in band n3 of the 5G NR specifications, intermodulation (IM) products leaking to the receiver (RX) may cause RX desensitization [2]. The intermodulation products typically stem from nonlinear devices after the transmitter (TX) chain, where they are unaffected by band-limiting filtering, and thus the non-contiguously allocated CCs can generate IM products to the RX chain. It is possible that the generated IM distortion stems from a passive device, such as switches, filters, or loosely connected cables. In these cases, the IM distortion is referred to as passive intermodulation (PIM) distortion. Additionally, the PIM distortion may originate completely from outside the transceiver chain, where e.g., rusted bolts, fences, or metallic rooftops may be source of the PIM. In such cases, we have termed the distortion air-induced PIM.

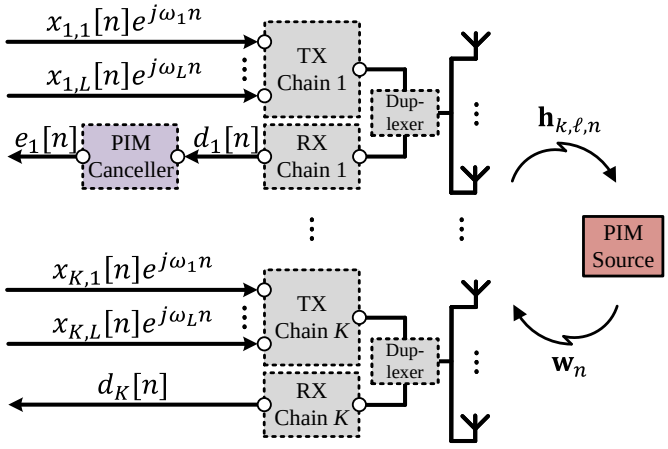

Fig. 1. Overall system model, where K TX chains transmit L concurrent carriers. A PIM source in the vicinity of the transceiver system distorts and reflects the TX signals back tot the RX, where a PIM canceller suppresses the received self-interference.

Arguably the simplest solution to the RX desensitization issue due to PIM self-interference (SI) is to reduce the TX signal power by backing off the TX power amplifiers (PAs), thereby also lowering the generated distortion signal power. However, this is not advisable as it leads to diminished power efficiency and coverage of the system. An alternative approach is to generate an estimate of the SI signal and subtract it from the RX data. While this identification and mitigation of the PIM distortion can be carried out also in the analog domain [3], in this work we focus on methods applied in the digital domain. Previous works e.g., in [4]–[6] have all demonstrated the feasibility of the digital domain PIM SI cancellers with conducted PIM sources. However, these works have limited the investigation to cases with a single TX chain, and therefore MIMO operation is omitted completely.

This paper builds upon our previous works [7]–[9] where air-induced PIM has been considered in a simple four carrier MIMO scenario. The works in [7], [8] introduced methods to reduce the overall computational complexity of the canceller engine, however, they are only applicable to the specific case of four carriers, such that two carriers are transmitted from two separate TX chains. While the method shown in [9] was reported to be able to perform in more evolved scenarios, measurements were only conducted in the aforementioned four carrier scenario, in an isolated chamber. Thus, to fill the

gap left by the previous works, in this paper we examine MIMO scenarios with more than four concurrent carriers, and also consider a case with channels under memory effects. We will develop an effective Hammerstein-type system, where the nonlinear effect of the PIM source is modeled with polynomial basis functions, and compare that against a slightly modified version of the Wiener-Hammerstein-type model introduced in [9]. From the conducted simulations, we can see that both models are capable of cancelling the received PIM signal close to the noise floor in the evolved scenarios when single tap channels are considered. Moreover, a scenario with channels with memory effects is simulated, where the Wiener-Hammerstein-type model is shown to benefit from the possibility to incorporate memory to the channel models.

II. System Model for Air-Induced PIM

In this section, we introduce a general model for a system where a BS operating in FDD mode transmits arbitrarily many signals, using arbitrarily many parallel TX chains simultaneously. These signals propagate to a PIM source through a wireless channel where they undergo nonlinear transformation. The nonlinear signal then propagates through another wireless channel back to the receiver, thus causing strong interference at the receiver.

Let us denote each baseband TX signal at time instance n as $x_{k,\ell}[n]e^{j\omega_\ell n}$, where $k = 1, 2, \cdots, K$ denotes the TX chain index with the total number of chains being K, $\ell = 1, 2, \cdots, L$ denotes the index of the signal in the TX chain, with L being the total number of signals in each chain, and ω_ℓ is the baseband center frequency of the ℓ-th carrier. Let the carriers be distributed symmetrically around zero in frequency, and let the frequency difference between consecutive carriers be $\omega_{\ell+1} - \omega_\ell = 2\omega_0$ for $\ell = 1, 2, \cdots, L-1$. Then, for the ℓ-th carrier, the center frequency can be defined as $\omega_\ell = (2\ell - 1 - L)\omega_0$. Now, the summed, total signal at the PIM source input $z[n]$ can be written as

$$z[n] = \sum_{k=1}^{K} \sum_{\ell=1}^{L} \mathbf{h}_{k,\ell,n}^T \mathbf{x}_{k,\ell,n}, \tag{1}$$

where $\mathbf{h}_{k,\ell,n} \in \mathbb{C}^{H\times 1}$ is the channel response filter associated with the ℓ-th carrier of the k-th TX chain, $\mathbf{x}_{k,\ell,n} \in \mathbb{C}^{H\times 1}$ collects H values of $x_{k,\ell}[n]e^{j\omega_\ell n}$, and $(\cdot)^T$ denotes the transpose operation. Additionally, H denotes the length of the filter $\mathbf{h}_{k,\ell,n}$, given by $H = H_1 + H_2 + 1$, where H_1 is the number of pre-cursor, and H_2 the number of post-cursor taps.

After the summation, the signal $z[n]$ then undergoes a nonlinear transformation in the PIM source. In this work, we model the nonlinear response with a standard polynomial model [10]. The instantaneous output of the PIM source $s[n]$, after the polynomial model is then given as

$$s[n] = \sum_{\substack{p=1 \\ p \text{ odd}}}^{P} g_p z[n]|z[n]|^{p-1}, \tag{2}$$

where P is the order of the model and g_p the p-th coefficient of the model.

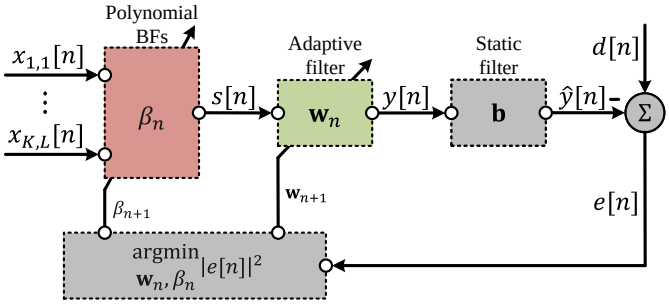

Fig. 2. Block diagram of the Hammerstein polynomial-based model.

After the nonlinear transformation, the distorted signal then propagates back to the receiver. Denoting the channel response with $\mathbf{w}_n \in \mathbb{C}^{M\times 1}$, the received signal $y[n]$ can be modeled as

$$y[n] = \mathbf{w}_n^T \mathbf{s}_n, \tag{3}$$

where $M = M_1 + M_2 + 1$ is the length of the filter $\mathbf{w}_n$, while M_1 and M_2 denote the pre- and post-cursor taps, respectively, and $\mathbf{s}_n$ collects M values of $s[n]$. The taps of $\mathbf{w}_n$ also capture some of the possible memory effects stemming from the nonlinearity. Finally, the received signal $y[n]$ is filtered in a band filter $\mathbf{b} \in \mathbb{R}^{B\times 1}$, with length $B = B_1 + B_2 + 1$, and B_1 pre-cursor and B_2 post-cursor taps, to produce a band-limited received signal $\hat{y}[n]$:

$$\hat{y}[n] = \mathbf{b}^T \mathbf{y}_n, \tag{4}$$

where $\mathbf{y}_n$ collects B samples of $y[n]$.

III. Cascaded Models for Cancellation of Air-Induced PIM

Stemming from the system model introduced in the previous section, we determine two possible realizations for a canceller modeling and cancelling the received air-PIM. The first will feature a typical Hammerstein model, where the nonlinearity is modeled using a vector of instantaneous polynomial basis functions (BFs). The latter is a slightly modified version of a technique introduced in [9], where a Wiener-Hammerstein-type cascaded model is utilized, accompanied by a spline-interpolated lookup table (LUT) for nonlinearity modeling. We will also show low-complexity gradient descent-based solutions to the parameter adaptation in both cases.

A. Hammerstein Polynomial-based Approach

The Hammerstein polynomial-based (HP) method is based on the assumption that the channels between the transceiver and the PIM source can be approximated by a single tap, i.e., $H = M = 1$. This approximation has been used also in our previous works [7], [8], where it has been shown to be a valid assumption in measured cases. Thus, the cancellation engine can be described as in Fig. 2, where we first have an instantaneous nonlinear transformation using polynomial BFs, followed by a global filter.

The instantaneous BFs required in the nonlinear block is dependent on the number of TX chains K, the number of signals at each chain L, as well as the RX operating frequency.

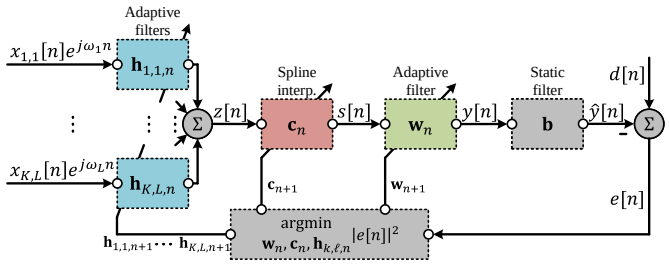

Fig. 3. Block diagram of the Wiener-Hammerstein LUT-based model.

Using the aforementioned single-tap approximation, the PIM source input $\tilde{z}[n]$ can be estimated as

$$\begin{aligned}\tilde{z}[n] &= \sum_{k=1}^{K}\sum_{\ell=1}^{L} h_{k,\ell}[n]x_{k,\ell}[n]e^{j\omega_\ell n} \\ &= \sum_{\ell=1}^{L} \psi_\ell[n]e^{j(2\ell-1-L)\omega_0 n},\end{aligned} \tag{5}$$

where $h_{k,\ell}[n]$ denotes the single tap channel coefficient associated with the ℓ-th carrier of the k-th TX chain, and $\psi_\ell[n]$ is the total summed signal at frequency ω_ℓ, over all TX chains.

In addition, the number of BFs is dependent on the polynomial order P considered, which we set to $P = 5$ in this paper. Then, following (2), we can write the approximated nonlinearity output $\tilde{s}[n]$ as

$$\tilde{s}[n] = g_1\tilde{z}[n] + g_3\tilde{z}[n]^2\tilde{z}^*[n] + g_5\tilde{z}[n]^3\tilde{z}^*[n]^2, \tag{6}$$

which can be expanded using (5) to a form as shown in (7) at the top of the next page. With (7), we can identify the required fifth order BFs at the wanted intermodulation (IM) frequency. Note that by considering single tap channels $h_{k,\ell}[n]$ per carrier, each BF will have a single effective coefficient, which lumps the channel coefficients and the nonlinearity coefficients g.

The obtained instantaneous BFs are collected to the column vector $\mathbf{\Phi}_n \in \mathbb{C}^{N\times 1}$, where N is the cardinality, i.e., number of BFs in the vector, where each carrier is taken at time instance n. To obtain the nonlinear block output, we need to weight each BF with its corresponding coefficient, which are collected to the vector $\beta_n \in \mathbb{C}^{N\times 1}$. Then, the instantaneous nonlinearity output is determined as

$$s[n] = \beta_n^T \mathbf{\Phi}_n, \tag{8}$$

after which the global filter $\mathbf{w}_n$ is used per (3) to consider potential memory effects in the nonlinear block, and the channel from the PIM source to the receiver. Lastly, the signal is filtered in the static low-pass filter $\mathbf{b}$ to limit the frequency range.

In order to track the coefficients of the nonlinear block and the filter $\mathbf{w}_n$, gradient descent-based algorithms are developed, where the coefficient updates are made in the negative direction of the gradient of the cost function $J[n] = |e[n]|^2$ [11]. Since the cancellation engine after the filter $\mathbf{w}_n$ is identical to the one presented in [9], we have

$$\begin{aligned}\mathbf{w}_{n+1} &= \mathbf{w}_n - \mu_w \frac{\partial J[n]}{\partial \mathbf{w}_n^*} \\ &= \mathbf{w}_n + \mu_w e[n]\mathbf{S}_n^*\mathbf{b},\end{aligned} \tag{9}$$

where μ is the step-size, and $\mathbf{S}_n \in \mathbb{C}^{M\times B}$ collects B past vectors $\mathbf{s}_n$. Similarly, we can determine the update rule for the BF coefficient vector to be

$$\begin{aligned}\beta_{n+1} &= \mathbf{w}_n - \mu_\beta \frac{\partial J[n]}{\partial \beta_n^*} \\ &= \beta_n + \mu_\beta e[n]\mathbf{\Gamma}_n^*\mathbf{b},\end{aligned} \tag{10}$$

where $\mathbf{\Gamma}_n = [\mathbf{\Omega}_n\mathbf{w}_n \ \cdots \ \mathbf{\Omega}_{n+B}\mathbf{w}_{n+B}] \in \mathbb{C}^{N\times B}$ collects B past vectors $\mathbf{\Omega}_n\mathbf{w}_n$, and where $\mathbf{\Omega}_n = [\mathbf{\Phi}_n \ \cdots \ \mathbf{\Phi}_{n+M}] \in \mathbb{C}^{N\times M}$ collects M past vectors $\mathbf{\Phi}_n$.

A well-known issue with the use of polynomial BFs in conjunction with gradient-based learning algorithms is the slow convergence due to the high correlation between the BFs. To alleviate this, the BFs can be multiplied with the inverse of the cross-correlation matrix $\mathbf{R}_n = \mathbb{E}[\mathbf{\Phi}_n\mathbf{\Phi}_n^H] \in \mathbb{C}^{N\times N}$, where $\mathbb{E}[\cdot]$ denotes the statistical expectation operator [6]. Since the matrix inversion is a resource hungry operation, we can save resources by precomputing the inverted cross-correlation matrix $\mathbf{R}^{-1}$ since the statistical properties of the BFs do not change over time. Then, we can use $\mathbf{R}^{-1}$ in the update of the BF coefficient vector as

$$\beta_{n+1} = \beta_n + \mu_\beta e[n](\mathbf{R}^*)^{-1}\mathbf{\Gamma}_n^*\mathbf{b}. \tag{11}$$

B. Wiener-Hammerstein LUT-based Approach

The overall principle of the Wiener-Hammerstein LUT-based (WHLUT) method was already established in [9], but is reviewed here briefly for convenience. Further, we will modify the technique from [9] slightly, by considering separate filters for each carrier signal, as described in the system model in Section II. This way, we can loosen the length requirements of the filters in certain cases, e.g., with 20 MHz BW in [9]. We also omit the carrier information from the LUT output, which is redundant when considering the IM products. The considered cancellation model is depicted in Fig. 3.

In the WHLUT method, the individual carriers $x_{k,\ell}[n]$ are first filtered in filters $\mathbf{h}_{k,\ell,n}$ and summed together to produce the total PIM source input $z[n]$, as per (1). As mentioned above, this is the difference to the algorithm introduced in [9], where the filtering was done to the aggregated signals from each TX chain. After the PIM source input is computed, the nonlinearity in the PIM source is modeled using a complex valued spline-interpolated LUT, with unit spaced knots. The PIM source output – without the redundant carrier information – can then be written as

$$s[n] = z[n]\mathbf{\Psi}_n^T\mathbf{c}_n, \tag{12}$$

$$\tilde{s}[n] = g_1 \sum_{\ell_1=1}^{L} \psi_{\ell_1}[n] e^{j(2\ell_1-1-L)\omega_0 n} + g_3 \sum_{\ell_1=1}^{L}\sum_{\ell_2=1}^{L}\sum_{\ell_3=1}^{L} \psi_{\ell_1}[n]\psi_{\ell_2}[n]\psi^*_{\ell_3}[n] e^{j(2\ell_1+2\ell_2-2\ell_3-1-L)\omega_0 n} + g_5 \sum_{\ell_1=1}^{L}\sum_{\ell_2=1}^{L}\sum_{\ell_3=1}^{L}\sum_{\ell_4=1}^{L}\sum_{\ell_5=1}^{L} \psi_{\ell_1}[n]\psi_{\ell_2}[n]\psi_{\ell_3}[n]\psi^*_{\ell_4}[n]\psi^*_{\ell_5}[n] e^{j(2\ell_1+2\ell_2+2\ell_3-2\ell_4-2\ell_5-1-L)\omega_0 n} \tag{7}$$

TABLE I
COMPUTATIONAL COMPLEXITIES OF THE INTRODUCED CASCADED MODELS IN TERMS OF FLOPS PER PROCESSED SAMPLE

Operation	**Complexity (FLOPs/sample)**	
	HP	**WHLUT**
Main Path	$G_{bf} + 10M + 10N - 4$	$KL(10H-2) + 2P_s^2 + 8P_s + 10M + 22$
Parameter Adaptation	$10N^2 + 4\tau(M+N) + 10MN - 3N + 20$	$2P_s^2 + P_s + 4\tau(M+C) + 2MC + 10M + 21$ $+KL(10HM + 4H\tau + 8M - 2H + 10)$

where $\mathbf{c}_n \in \mathbb{C}^{C\times 1}$ is the spline control point vector with C elements, while $\mathbf{\Psi}_n \in \mathbb{R}^{C\times 1}$ is the spline basis function vector, given as

$$\mathbf{\Psi}_n = [0 \ \cdots \ 0 \ \mathbf{u}_n^T \mathbf{C}_{P_s} \ 0 \ \cdots \ 0]^T, \tag{13}$$

where $\mathbf{C}_{P_s} \in \mathbb{C}^{(P_s+1)\times(P_s+1)}$ denotes a static coefficient matrix, defined separately for each interpolation order P_s. The so-called abscissa vector $\mathbf{u}_n \in \mathbb{R}^{1\times(P_s+1)}$ is defined as

$$\mathbf{u}_n = [u[n]^{P_s} \ u[n]^{P_s-1} \ \cdots \ 1], \tag{14}$$

where $u[n] = |z[n]| - (i[n]-1)$, and $i[n] = \lfloor |z[n]| \rfloor + 1$, when considering unit spaced knots. In (13), the non-zero elements start at index $i[n]$. After the nonlinear transformation according to (12), the global filter is applied after (3), and the band is limited after (4).

Lastly, the gradient descent-based update rules for the WHLUT coefficients are presented. For the global filter $\mathbf{w}_n$, the update rule is the same as in (9). The learning rule for the spline control point vector is given as

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \mu_c e[n] \mathbf{\Pi}_n^* \mathbf{b}, \tag{15}$$

where the columns of $\mathbf{\Pi}_n \in \mathbb{C}^{C\times B}$ are delayed versions of $\mathbf{\Sigma}_n \mathbf{Z}_n \mathbf{w}_n$, where $\mathbf{\Sigma}_n = [\mathbf{\Psi}_n \ \cdots \ \mathbf{\Psi}_{n+M}]$, and $\mathbf{Z}_n = diag\{z[n] \ \cdots \ z[n+M]\}$. The update for each filter $\mathbf{h}_{k,\ell,n}$ is given as

$$\mathbf{h}_{k,\ell,n+1} = \mathbf{h}_{k,\ell,n} + \mu_h e[n] \mathbf{\Xi}^*_{k,\ell,n} \mathbf{b}, \tag{16}$$

where the columns of the matrix $\mathbf{\Xi}_{k,\ell,n} \in \mathbb{C}^{H\times B}$ are delayed copies of $\mathbf{\Upsilon}_{k,\ell,n}\mathbf{w}_n$, where the columns of $\mathbf{\Upsilon}_{k,\ell,n} \in \mathbb{C}^{H\times M}$ in turn are delayed versions of $\mathbf{x}_{k,\ell,n}\delta[n]$. The coefficient $\delta[n]$ is given as

$$\delta[n] = \frac{|z[n]|}{2} \dot{\mathbf{u}}_n^T \mathbf{C}_{P_s} \tilde{\mathbf{c}}_n + \mathbf{\Psi}_n^T \mathbf{c}_n, \tag{17}$$

where $\dot{\mathbf{u}}_n = \left[P_s u[n]^{P_s-1} \ (P_s-1)u[n]^{P_s-2} \ \cdots \ 1 \ 0\right]^T \in \mathbb{R}^{(P_s+1)\times 1}$ contains the derivatives of $u[n]^{P_s}$, and $\tilde{\mathbf{c}}_n$ contains $P+1$ elements of $\mathbf{c}_n$ starting from index $i[n]$.

C. Computational Complexity

The computational complexities of the HP and WHLUT methods are presented in Table I in terms of floating point operations (FLOPs) per processed sample. It is assumed here that a multiplication between complex numbers takes 8 FLOPs, multiplication between a complex and a real number takes 2 FLOPs, addition of complex numbers takes 2 FLOPs, and addition of a complex and a real number takes a single FLOP. The complexity of filtering with the static filter $\mathbf{b}$ is omitted from Table I.

The number of BFs, as denoted by N, as well as the computational complexity associated with the generation of the BFs, denoted by G_{bf} in Table I, depend on the number of TX chains K and CCs per chain L. The total number of BFs can be computed by first determining the distinct configurations of ψ_ℓ at the wanted frequency. For example, the configurations for $L=2$, $K=3$, at $-(L+1)\omega_0$ are $\psi_1^2[n]\psi_2^*[n]$, $\psi_1^3[n]\psi_1^*[n]\psi_2^*[n]$, and $\psi_1^2[n]\psi_2[n]\psi_2^*[n]^2$. Each $\psi_\ell^p[n]$ generates $\binom{p+K-1}{K-1}$ distinct BFs, as per the number of multinomial coefficients. Thus, the aforementioned configurations generate $6\times 3 = 18$, $10\times 3\times 3 = 90$, and $6\times 3\times 6 = 108$ BFs, respectively, for a total of 216 BFs. For G_{bf}, it is assumed that intermediate results can be reused to lower the total complexity. As an example, it takes 536 FLOPs per sample to generate the required BFs when $L=K=2$. The number of BFs N and the complexities to generate the BF G_{bf} for various configurations of K and L are shown in in Table II.

The learning complexity can also be brought down if only $\tau < B$ most significant taps around the center tap of the filter $\mathbf{b}$ are considered. This has been previously shown to yield excellent results, with much reduced complexity [9]. We will show numerical results for the complexities with various parametrizations of the system in the next section.

IV. SIMULATIONS AND RESULTS

In this section, we will verify the operation of the cascaded models introduced in the previous section by simulations. The

TABLE II

ACHIEVED CANCELLATION RESULTS AND ASSOCIATED COMPUTATIONAL COMPLEXITIES WITH VARIOUS NUMBER OF TX CHAINS K AND CCS PER CHAIN L, CONSIDERING SINGLE TAP CHANNELS AND 5 MHZ BW CARRIERS.

K	L	Cancellation (dB)			N	G_{bf}	FLOPs/sample	
		HP	WHLUT	Maximum			HP	WHLUT
2	2	20.3	20.2	20.6	40	536	942 / 21,680 / 22,622	351 / 4,374 / 4,725
	3	20.4	19.9	20.6	156	2,352	3,958 / 264,236 / 268,194	447 / 5,794 / 6,241
3	2	20.4	20.2	20.6	216	2,432	4,638 / 495,296 / 499,934	447 / 5,794 / 6,241
	3	20.2	19.9	20.5	951	12,488	22,044 / 9,169,031 / 9,191,075	573 / 10,177 / 10,750

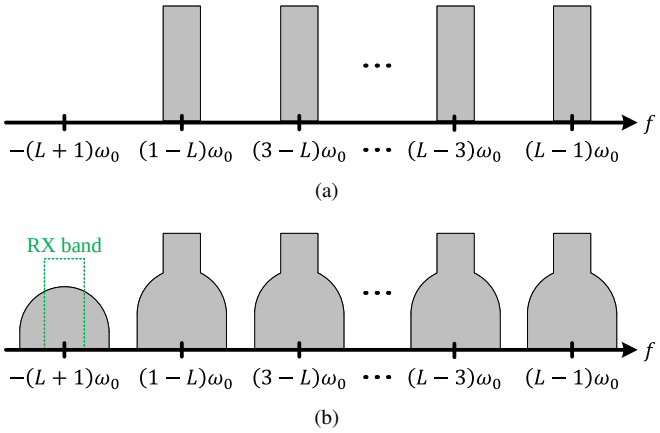

Fig. 4. Concept of the spectra of the signals (a) before and (b) after the nonlinear transformation in the PIM source.

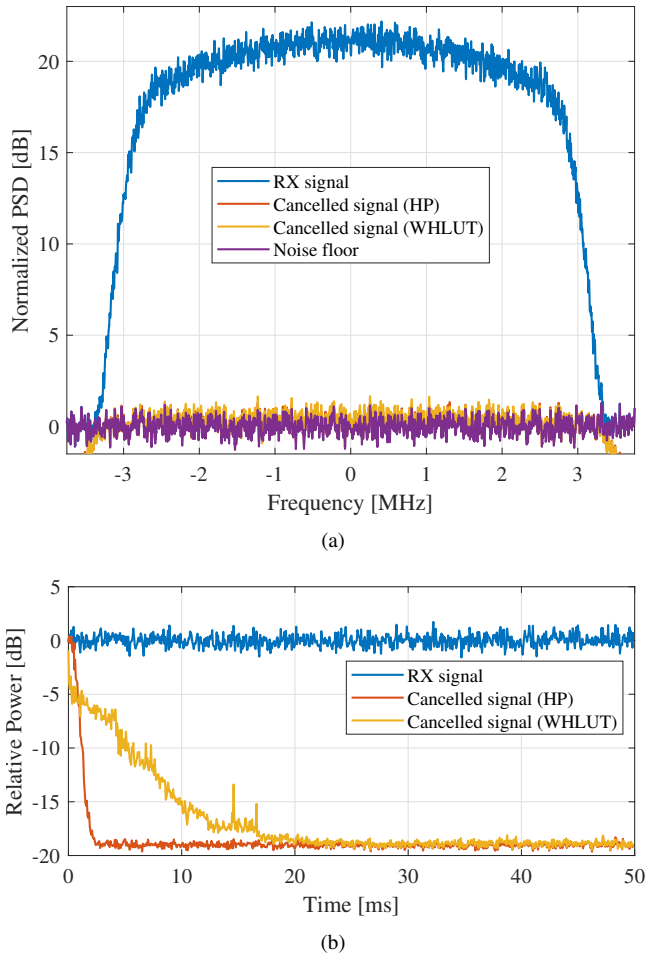

Fig. 5. Spectrum (a) and convergence curve (b) of the signals involved in the cancellation with both considered methods, in the case $K = L = 2$, single tap channels.

system model is based on the one introduced in Section II, and we will vary the number of TX chains K, CCs per chain L, and the memory in the channels H and M between the scenarios. Here, we concentrate on PIM distortion at $\omega_{RX} = -(L+1)\omega_0$, which corresponds to the closest negative third order IM band, i.e., the band $2\omega_0$ below the first signal band $(1-L)\omega_0$, as illustrated in Fig. 4. In order to utilize the HP method we have to separately define the BFs for each configuration of K and L. To find the BFs at $-(L+1)\omega_0$, we need to set $\ell_1+\ell_2 = \ell_3$ for the third order BFs, and $\ell_1+\ell_2+\ell_3 = \ell_4+\ell_5$ for the fifth order BFs in (7). The generated TX signals are 5G NR compliant OFDM waveforms with 5 MHz bandwidth, and with around 7 dB peak-to-average power ratio (PAPR). The signals are generated at the fundamental frequency of 7.68 MHz, and are oversampled by a factor of 16 for the purposes of the cancellers. The utilized polynomial model coefficients g_p in (2) in the system model are from a PA measurement, where least squares fitting was used to identify the coefficients.

First, we will investigate the cancellation fidelity of the introduced methods with single tap channels, which – as mentioned previously – is a valid approximation in many air-induced PIM scenarios. For both cancellers, we have set $M = 5$ and $\tau = 21$ and in the WHLUT method $H = 5$, $C = 11$, and $P_s = 3$. Table II shows the cancellation results with four distinct scenarios, with two and three TX chains and with two and three simultaneous CCs. In each of the scenarios, the maximum available cancellation is around 21 dB, which the cancellers are almost able to achieve, with only minor residual powers from the received PIM signals. A snapshot of the signal spectra as well as the convergence curves of the first scenario, i.e., where $K = L = 2$ are shown in Fig. 5. From the convergence curve, it is discernible that the convergence of the HP model is extremely fast, taking only around 2 ms, while the full convergence of the WHLUT model takes around 20 ms. Additionally, the computational complexity of both models to in each scenario is presented in Table II. Here, we see

that to achieve the high cancellation fidelity in the HP model, the computational complexity increases rapidly with increased number of TX chain and/or CCs. Meanwhile, the computational complexity associated with the WHLUT method also increases with more CCs, however, the increase is not as drastic, and we can see from Table II that only the total number of CCs affects the complexity, not how they are assigned to the TX chains. It is important to note here that the oversampling requirement of the HP method is not as stringent as in the WHLUT method. This stems from the WHLUT method requiring the products on the carrier frequencies as well as the IM product information in the main path, while the HP method operates solely on the BFs of the IM product in question. Yet, the complexities of the cases with more than four concurrent CCs (i.e., when $K > 2$ and/or $L > 2$) are so high that even with reduced sampling rate requirement the HP method still requires more computations per time unit than the WHLUT method. We can therefore conclude that both methods are able to model and cancel the received PIM with high accuracy when single tap channels are considered, with the HP method requiring less time for convergence. This reduced convergence time is achieved with high computational complexity, which rapidly rises with increased parameter count. On the other hand, the slightly slower convergence time of the WHLUT method is achieved with much reduced complexity.

Lastly, let us consider the case $K = L = 2$ under channels with memory effects. Fig. 6 presents the signals spectra and convergence curves for both canceller models when each channel has 3 pre- and post-cursor taps, while the center tap is still the most significant. Due to our memoryless channel assumption in the HP method, the achieved cancellation level is reduced compared to the WHLUT method, where the channels' memory effects can be modeled. It can be seen that the cancellation achieved by the WHLUT method almost reaches the noise floor, the achieved cancellation being around 19.9 dB out of 20.5 dB maximum. Meanwhile, the cancellation of the HP method is around 15 dB. Compared to the memoryless channel case, the convergence time is now much longer: it takes around 400 ms to reach convergence with the WHLUT method, although a level of around 12 dB is reached already at 8 ms.

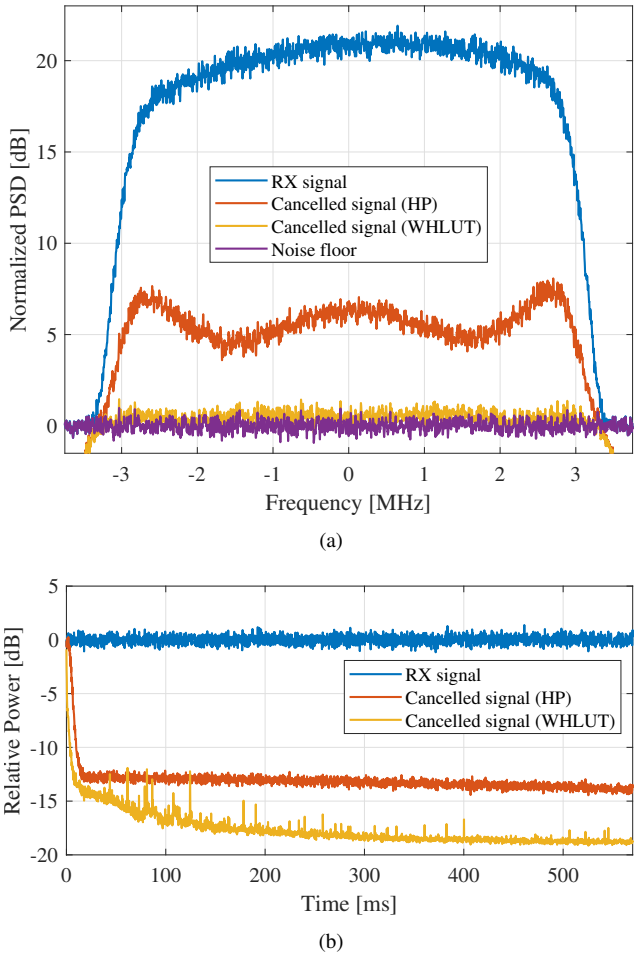

Fig. 6. Spectrum (a) and convergence curve (b) of the signals involved in the cancellation with both considered methods, in the case $K = L = 2$, under channels with memory effects.

V. Conclusion

This paper has investigated air-induced PIM in generic MIMO scenarios. Unlike prior works, we allowed there to co-exist any number of TX chains and CCs in each TX chain. Leveraging on the assumption of single tap channels, we introduced the Hammerstein polynomial model, where the wireless channels and the nonlinear coefficients are lumped, while a global filter after the nonlinear block models both the potential memory of the nonlinear device and the wireless channel back to the receiver. As a reference, we also showed a slightly modified version of the Wiener-Hammerstein-type LUT-based model, introduced in our earlier work. The measurements and complexity assessment show that while the Hammerstein polynomial model showcases excellent SI cancellation, the cost in computations rises rapidly with the total number of CCs in the system. On the other hand, the Wiener-Hammerstein LUT-based method can operate with much reduced complexity, and can achieve practically the same performance levels as the polynomial-based model. In addition, the LUT-model can cancel the SI when the SI channels have memory effects, while the polynomial Hammerstein model suffers from the single tap channel approximation.

Acknowledgement

This work was supported in part by Nokia Mobile Networks, and in part by the Academy of Finland under Grants 323461, 321613, and 301820.

References

[1] 3GPP Tech. Spec. 38.104, "NR; Base Station (BS) radio transmission and reception," *v18.2.0, (Release 18)*, June 2023.

[2] A. Kiayani, V. Lehtinen, L. Anttila, T. Lahteensuo, and M. Valkama, "Linearity Challenges of LTE-Advanced Mobile Transmitters: Requirements and Potential Solutions," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 170–179, 2017.

[3] J. J. Henrie, A. J. Christianson, and W. J. Chappell, "Linear–Nonlinear Interaction and Passive Intermodulation Distortion," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 5, pp. 1230–1237, 2010.

[4] H.-T. Dabag, H. Gheidi, S. Farsi, P. S. Gudem, and P. M. Asbeck, "All-Digital Cancellation Technique to Mitigate Receiver Desensitization in Uplink Carrier Aggregation in Cellular Handsets," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 12, pp. 4754–4765, 2013.

[5] H. Gheidi, H.-T. Dabag, Y. Liu, P. M. Asbeck, and P. Gudem, "Digital cancellation technique to mitigate receiver desensitization in cellular handsets operating in carrier aggregation mode with multiple uplinks and multiple downlinks," in *2015 IEEE Radio and Wireless Symposium (RWS)*, 2015, pp. 221–224.

[6] M. Z. Waheed, D. Korpi, L. Anttila, A. Kiayani, M. Kosunen, K. Stadius, P. P. Campo, M. Turunen, M. Allén, J. Ryynänen, and M. Valkama, "Passive Intermodulation in Simultaneous Transmit–Receive Systems: Modeling and Digital Cancellation Methods," *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 9, pp. 3633–3652, 2020.

[7] V. Lampu, L. Anttila, M. Turunen, M. Fleischer, J. Hellmann, and M. Valkama, "Air-Induced PIM Cancellation in FDD MIMO Transceivers," *IEEE Microwave and Wireless Components Letters*, vol. 32, no. 6, pp. 780–783, 2022.

[8] ——, "Air-Induced Passive Intermodulation in FDD MIMO Systems: Algorithms and Measurements," *IEEE Transactions on Microwave Theory and Techniques*, vol. 71, no. 1, pp. 373–388, 2023.

[9] ——, "Cancellation of Air-Induced Passive Intermodulation in FDD MIMO Systems: Low-Complexity Cascade Model and Measurements," in *Proc. 2023 IEEE/MTT-S International Microwave Symposium - IMS 2023*, 2023, pp. 33–36.

[10] D. Morgan, Z. Ma, J. Kim, M. Zierdt, and J. Pastalan, "A Generalized Memory Polynomial Model for Digital Predistortion of RF Power Amplifiers," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3852–3860, Oct. 2006.

[11] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Prentice Hall, 1996.