SCALE-X: A Systematic Complexity-Aware Low-Precision Approach for Digital Predistortion

Anees Rehman, Mohd Tasleem Khan, *Senior Member, IEEE,* George Goussetis, *Fellow, IEEE,* Yuan Ding, *Member, IEEE,* Jaiyu Hou, and João F. C. Mota, *Member, IEEE*

***Abstract*—This letter presents the SCALE-X framework for digital predistortion, a systematic, complexity-aware, model-based approach. The proposed scheme employs a novel complexity-aware coefficient pruning method, which, when combined with optimized fixed-point modeling, enables accurate capture of quantization effects and supports better-informed decisions for implementation. Targeted configuration of model order, memory depth, and coefficient bitwidth dramatically reduces computational overhead without noticeable reduction in accuracy. Field-programmable gate array implementation demonstrates practical trade-offs with improved efficiency.**

***Index Terms*—Digital Pre-distortion (DPD), Field programmable gate array (FPGA), Pruning, Quantization**

I. Introduction

POWER amplifiers (PAs) are inherently nonlinear devices, introducing spectral regrowth and signal distortion that degrade the performance of modern wireless communication systems [1]. Digital predistortion (DPD) is the predominant linearization technique, suppressing out-of-band emissions and restoring signal fidelity without sacrificing PA efficiency [2]. With more stringent requirements in beyond-5G, DPD must reduce the adjacent channel power ratio (ACPR) and error vector magnitude (EVM) targets — 3GPP 5G NR mandates $\geq$ 45 dBc and 3.5–8% respectively [3]–[5]. However, deploying DPD on resource-constrained platforms such as field-programmable gate arrays (FPGAs) makes it challenging to maintain high-fidelity linearization within tight hardware budgets [6]–[8]. Understanding how DPD model complexity scales with hardware budget is therefore essential.

Model-based approaches such as memory polynomial (MP) [9], dynamic deviation reduction [10], and Volterra models [5] achieve accurate linearization by capturing nonlinearity and memory effects [11]. However, increasing polynomial order, or memory depth, or both increases the number of coefficients, and thus the computational cost [9], [12]. Among model-based techniques, MP-DPD offers a favorable trade-off between linearization performance and implementation complexity, making it widely adopted [13]. Its coefficients are typically estimated using least-squares (LS) fitting [14], which sometimes can yield ill-conditioned values in over-parameterized models [15]. Unlike LS, constrained LS (CLS) bounds coefficient magnitudes to improve numerical stability [16]. However, the impact of CLS on fixed-point (FxP) implementation [17] of DPD remains largely unstudied [18].

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/Y037197/1.

Model pruning [19] and term selection [20] retain only the most significant basis functions, reducing the number of multiplication operations while preserving linearization performance. Alternative basis functions, such as splines [21] and decomposed vector rotations [22], together with look-up table approaches [15], further cut computational overhead, while real-time model switching adapts active terms per sample to reduce complexity [23]. On the implementation side, FPGA realizations achieve efficient DPD within tight resource and energy budgets [6], [11], and efficient multiplier designs reduce hardware cost, often at the expense of latency and accuracy [24], [25]. Yet a unified treatment of FxP quantization and coefficient pruning remains absent from the literature. To bridge this gap, we propose SCALE-X, a complexity-aware framework for FPGA-based DPD. The main contributions are:

- We propose an anti-diagonal coefficient pruning strategy that reduces MP model complexity while preserving linearization performance.
- We develop a quantization-aware FxP model that jointly optimizes pruning and wordlength, enabling efficient accuracy–complexity trade-offs.
- We validate the proposed architecture on FPGA achieving reduced logic utilization and improved overall efficiency.

The remainder of this letter is organised as follows: Section II presents the SCALE-X framework; Section III benchmarks the design against the state-of-the-art; and Section IV concludes.

II. Proposed SCALE-X Framework

In model-based approaches, DPD is typically developed using the PA's inverse model with an indirect learning architecture [4]. Generally, the output $y(n)$ is expressed as

$$y(n) = \mathrm{sum}\big(\mathbf{X}(n) \circ \mathbf{C}\big), \tag{1}$$

where '$\circ$' denotes Hadamard elementwise multiplication, $\mathbf{X}(n) = [\chi(n-m,k)]_{M\times K}$, $\mathbf{C} = [c(m,k)]_{M\times K}$, $m = 0,\ldots,M-1$, $k = 1,\ldots,K$, and for the MP model,

$$\chi(n-m,k) = x(n-m)|x(n-m)|^{k-1}.$$

This requires $M(2K-1)$ multiplications (including the generation of nonlinearities) and $(MK-1)$ additions. Alternatively, (1) can be expressed as

$$y(n) = \mathrm{sum}\left(\boldsymbol{\chi}_0(n) \circ \left(\sum\nolimits_{k=1}^{K}\left(|\boldsymbol{\chi}_0(n)|^{\circ(k-1)}\right) \circ \mathbf{c}_k\right)\right), \tag{2}$$

where $\circ(k-1)$ denotes the Hadamard power, $\boldsymbol{\chi}_0(n) = \{x(n-m)\}_{m=0}^{M-1}$ is the input vector of memory depth M, and $\mathbf{c}_k =$

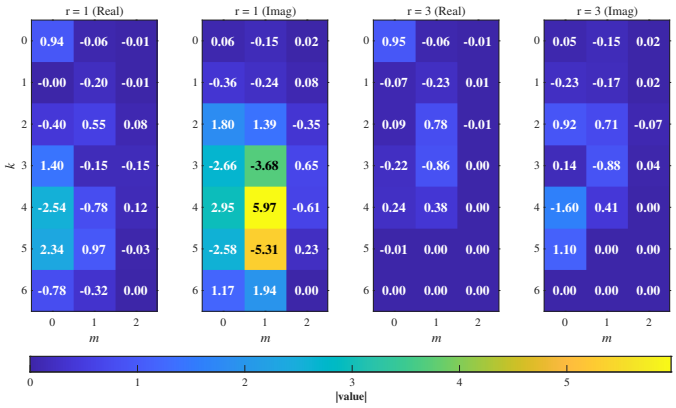

Fig. 1. Illustration of coefficients' heatmap for $K = 7$ and $M = 3$ with $r = 1$ (left) and $r = 3$ (right) to demonstrate the proposed approach.

$\{c_{mk}\}_{m=0}^{M-1}$ is the coefficient vector for nonlinearity order K. Unlike (1), (2) requires only MK multiplications.

The representation (2) reduces multiplications via Horner's rule [18], yet both formulations retain MK coefficients with identical normalized mean-square error (NMSE), ACPR, and EVM performance. This work therefore proposes a novel approach that selectively prunes the coefficient matrix $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_K] \in \mathbb{C}^{M\times K}$, reducing computational cost while preserving linearization performance. From simulation, it is observed that the coefficients toward the bottom-right corner of $\mathbf{C}$, corresponding to larger memory indices m and nonlinear orders k, contribute least to linearization. These coefficients are therefore removed progressively using an anti-diagonal pruning strategy that ranks coefficients by $m+k$. Specifically, with $m = 0, \ldots, M-1$ and $k = 1, \ldots, K$, the rth anti-diagonal to be pruned is defined as

$$\mathcal{D}_r = \{(m,k) : m + k = M + K - 1 - r\},$$

where $r = 0, 1, \ldots, (M+K-2)$ and $\mathcal{D}_0$ contains the bottom-right coefficient $c_{M-1,K}$. The cumulative set of pruned coefficients after stage r is $\mathcal{Z}_r = \bigcup_{d=0}^{r} \mathcal{D}_d$, and the corresponding active coefficient set is

$$\mathcal{A}_r = \{(m,k) : 0 \le m \le M-1,\ 1 \le k \le K\} \setminus \mathcal{Z}_r.$$

Thus, the pruned coefficient matrix is

$$[\mathbf{C}^{(r)}]_{mk} = \begin{cases} 0, & (m,k) \in \mathcal{Z}_r, \\ c_{mk}, & (m,k) \in \mathcal{A}_r. \end{cases}$$

At each pruning stage, the active coefficients are re-estimated using CLS, with coefficients in $\mathcal{Z}_r$ constrained to zero, rather than simply setting the corresponding coefficients of the original solution to zero. Specifically,

$$\hat{\mathbf{c}}_{\mathcal{A}_r}^{(r)} = \arg\min_{\mathbf{c}_{\mathcal{A}_r}} \|\mathbf{y} - \mathbf{\Phi}_{\mathcal{A}_r}\mathbf{c}_{\mathcal{A}_r}\|_2^2, \tag{3}$$

where $\mathbf{\Phi}_{\mathcal{A}_r}$ contains only the columns of the regression matrix associated with the active coefficient set $\mathcal{A}_r$. The remaining coefficients in $\mathcal{Z}_r$ are fixed to zero. The resulting $\hat{\mathbf{c}}^{(r)}$ is substituted into (2) to evaluate the linearised output, and the corresponding NMSE, ACPR, and EVM are recorded. Pruning continues until the first stage at which the linearization

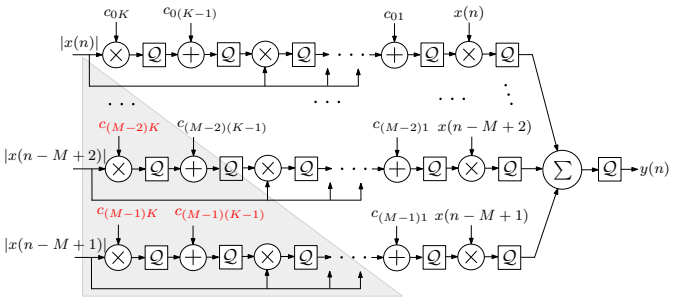

Fig. 2. Proposed architecture for (4); pruned coefficients are in shaded area.

performance falls outside the prescribed limits. Specifically, a pruning stage is considered feasible when NMSE and EVM remain below their respective limits and ACPR remains above its required minimum. If any of these performance requirements is violated, the previous stage is selected as the maximum feasible pruning stage, i.e., $r^* = r - 1$. The resulting $\mathbf{C}^{(r^*)}$ therefore provides the maximum pruning achieved by the proposed sequential strategy while satisfying the specified linearization requirements. An example of the proposed approach, with coefficient heatmaps for $K = 7$ and $M = 3$ at $r = 1$ and $r = 3$, based on the simulation settings in Section III, is shown in Fig. 1.

For FPGA realization, inputs, coefficients, and intermediate products must be represented using FxP arithmetic. To account for finite precision, (2) is extended by applying an FxP quantizer to each intermediate product as follows:

$$y(n) = \mathcal{Q}\left[\text{sum}\left(\mathcal{Q}\left[\boldsymbol{\chi}_0(n) \circ \left(\mathcal{Q}\left[\sum_{k=1}^{K} \mathcal{Q}[\mathbf{p}_k]\right]\right)\right]\right)\right], \tag{4}$$

where $\mathcal{Q}[\cdot]$ denotes FxP quantization in signed $\mathcal{Q}_{I.F}$ format, with wordlength $W = I + F$ bits comprising I integer bits (including sign), and F fractional bits. The parameters (I, F) are jointly optimized alongside the pruning stage r^*, minimizing hardware complexity while maintaining acceptable linearization performance.

The FPGA architecture implementing (4) is shown in Fig. 2. Each of the M rows corresponds to a delayed input sample $x(n-m)$, whose magnitude $|x(n-m)|$ is raised to successive powers to form the nonlinear basis across K orders, consistent with $\mathbf{p}_k$ in (2). Within each row, basis functions are multiplied by c_{mk} and accumulated across all K orders via a chain of multiplier, quantizer and adder stages, realizing the inner sum of (4); row outputs are then summed across all M taps to yield $y(n)$, with a final quantization stage at the output. Coefficients highlighted in the shaded area correspond to the pruned anti-diagonals $\mathcal{D}_0, \ldots, \mathcal{D}_{r^*}$ of $\mathbf{C}^{(r^*)}$, zeroed by the proposed method. Algorithm 1 summarizes the pruning strategy and FxP optimization of the SCALE-X framework.

III. Results and Discussion

Experiments were conducted using an in-house PA built around the Qorvo QPA9822 at 3.5 GHz, with a measured small-signal gain of 38.24 dB and an input P1dB of −11.23 dBm. A Xilinx ZC706 FPGA platform with an FMCOMMS5 radio board generated the PA input and captured the

Algorithm 1 Pruning and FxP Optimization

Input: $x(n), y(n)$, M, K; **Output:** $\mathbf{C}^{(r^*)}$, (I^*, F^*)
Initialize: $\mathbf{C}^{(-1)} \leftarrow$ CLS solution, $\mathcal{Z}_{-1} \leftarrow \emptyset$
for $r = 0$ to $M + K - 2$ **do**
  $\mathcal{D}_r \leftarrow \{(m,k) : m + k = M + K - 1 - r\}$;
  $\mathcal{Z}_r \leftarrow \mathcal{Z}_{r-1} \cup \mathcal{D}_r$; $\mathcal{A}_r \leftarrow \{(m,k)\} \setminus \mathcal{Z}_r$
  $\hat{\mathbf{c}}_{\mathcal{A}_r}^{(r)} \leftarrow \arg\min \|\mathbf{y} - \mathbf{\Phi}_{\mathcal{A}_r} \mathbf{c}_{\mathcal{A}_r}\|_2^2$, $\mathbf{C}^{(r)} \leftarrow [\hat{\mathbf{c}}_{\mathcal{A}_r}^{(r)}, \mathbf{0}_{\mathcal{Z}_r}]$
  Evaluate NMSE, ACPR, EVM;
  if constraint violated: $r^* \leftarrow r - 1$, **break**
end for
for each (I, F) **do**
  $\mathbf{C}_{I,F} \leftarrow \mathcal{Q}_{I,F}(\mathbf{C}^{(r^*)})$; check feasibility
end for
$(I^*, F^*) \leftarrow \arg\min_{(I,F) \text{ feasible}} I + F$
return $(\mathbf{C}^{(r^*)}, I^*, F^*)$

PA output, as shown in Fig. 3. The test signal was a 10 MHz 5G NR waveform with a PAPR of 9.3 dB (at 0.01% CCDF), sampled at 40 MSPS. The PA operated at 6.3 dB average input back-off from P1dB, with peaks compressed by 2.8 dB.

A. Performance Evaluation

Fig. 4 shows the trade-off between model complexity and bit precision on DPD linearization performance (ACPR, EVM, NMSE). The floating-point baseline and the proposed method, under identical conditions and an equal number of active coefficients, across three memory-depth configurations, $M = 2$, $M = 3$, and $M = 4$, each pruned progressively from $r = 0$ (full model) toward higher r (fewer coefficients). The FxP with bit-width shown on the top axis for direct comparison against the fixed-point curve. Among the three, $M = 3$ and $K = 7$ offers the best accuracy–complexity balance, sustaining near-floating-point ACPR (-59.63 dBc), EVM (0.088%), and NMSE (-57.63 dB) down to roughly 12–14 coefficients before all three metrics degrade sharply beyond that point. $M = 2$ saturates at a noticeably worse performance floor even with minimal pruning, showing that its shallow memory depth limits achievable linearization regardless of coefficient count, while $M = 4$ matches or exceeds $M = 3$ only at high coefficient counts and degrades sooner under pruning, reflecting redundancy among its additional memory terms.

The FxP curve, obtained by quantizing the pruned $M = 3$ coefficients from Q1.15 down to Q1.11, closely tracks the floating-point pruning trend across all three metrics down to roughly 12–13 bits, indicating that reducing precision over this range costs little beyond what pruning alone already costs. Below this point, quantization noise begins to dominate and the FxP curve diverges from the floating-point baseline more sharply than coefficient count alone would predict, marking the point where precision, rather than model complexity, becomes the binding constraint on performance. This crossover indicates a practical joint operating point, roughly 12–14 coefficients at 12–13 bits, where pruning and FxP quantization can be combined with negligible compounding loss, directly motivating the SCALE-X co-design strategy of optimizing complexity and precision together rather than sequentially.

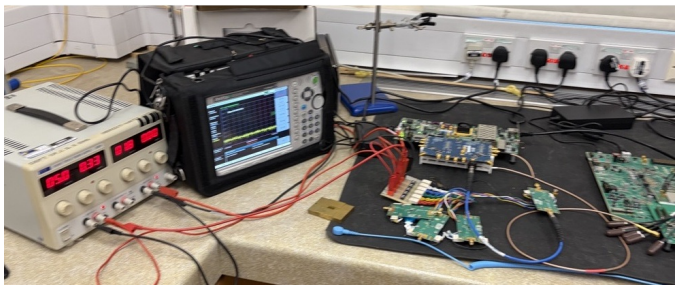

Fig. 3. Experimental setup, in which a ZC706 FPGA with an FMCOMMS5 transceiver drives and captures the QPA9822 PA at 3.5 GHz.

Fig. 5 compares the two proposed strategies, anti-diagonal coefficient pruning (blue) and direct $M \times K$ matrix-size reduction (orange), at matched coefficient counts across the following cases: $r = 2$ vs. 3×6 (left), $r = 3$ vs. 3×5, $r = 4$ vs. 3×4 and $r = 5$ vs. 3×3 (right), with the PA output without DPD (black) shown as reference. Both strategies substantially suppress adjacent-channel spectral regrowth relative to the no-DPD case, but anti-diagonal pruning consistently achieves lower out-of-band PSD than matrix-size reduction, particularly in the near-band region ($\pm$5–10 MHz), where the orange trace sits above the blue trace across all four sub-plots. This gap is largest at higher coefficient counts ($r = 2$ vs 3×6, $r = 3$ vs 3×5) and narrows as coefficients are reduced further ($r = 4$ vs 3×4, $r = 5$ vs 3×3), though anti-diagonal pruning retains a consistent edge even at the lowest complexity point. This confirms that, for a fixed coefficient count, selectively pruning the least significant basis functions preserves linearization performance better than uniformly truncating the model order, supporting the pruning strategy adopted in SCALE-X.

Table I reports the FPGA resource utilization and power consumption of the proposed method across pruning stages $r = 0$ to $r = 5$. The equivalent number of slices (ENS), defined in the table notes, condenses resource usage into a single metric. Its $\max$ term reflects that look-up tables (LUTs) and flip-flops (FFs) share the same slice, with weights corresponding to the soft-logic cost of replicating a 25×18 DSP multiply–accumulate and a 36 kb block RAM, respectively. A 7-series slice (4 LUT6, 8 FF) is adopted as the common unit for all designs, since native slice counts on UltraScale+ parts differ (8 LUT6, 16 FF). As the compared devices span 28 nm (Zynq-7000 7Z010) and 16 nm FinFET (ZU48DR, VU9P) process nodes with differing DSP primitives, ENS is read as an architecture-level comparison of logic quantity. Since sample rates differ across designs, ENS is normalized by throughput (TP), giving ENS/TP in slices per Msps.

As r increases, progressive zeroing of anti-diagonal entries in $\mathbf{C}^{(r)}$ directly reduces active multiplications: DSP usage falls from 50 to 25 (50% reduction), with LUTs and FFs following the same trend, from 1423 to 1168 and 718 to 471, respectively. This drives ENS down from 7856 at $r = 0$ to 4042 at $r = 5$ (48.6% reduction), almost entirely through DSP elimination on the pruned anti-diagonals. Since throughput remains constant at 375 Msps across all stages, ENS/TP falls by the same proportion, from 21.0 to 10.8. Power decreases correspondingly from 218 to 177 mW, confirming tangible en-

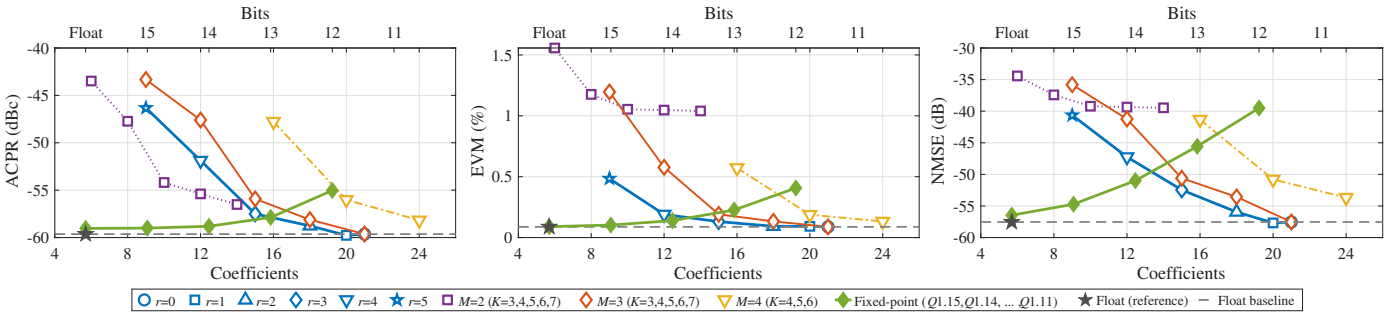

Fig. 4. Trade-off between model complexity and DPD linearization performance for three metrics: ACPR (left), EVM (middle), and NMSE (right); where dashed line corresponds to ideal linear output.

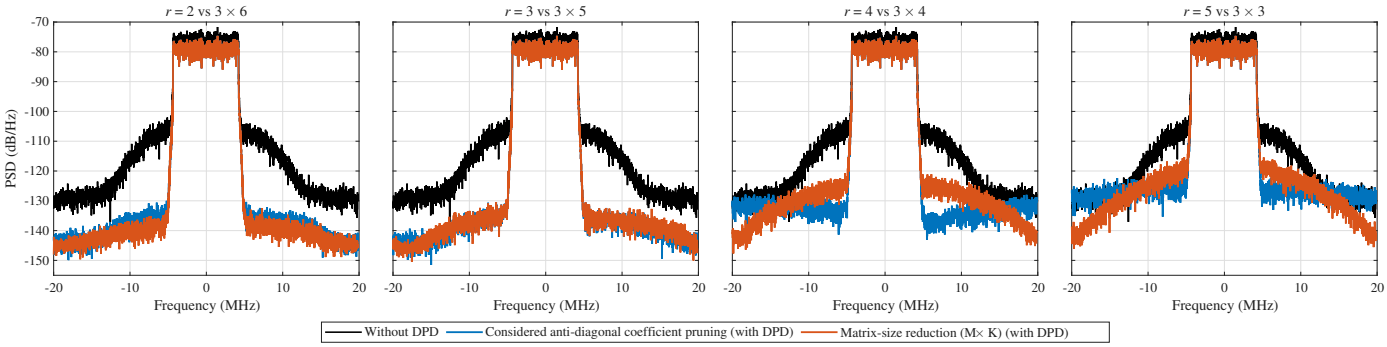

Fig. 5. Power spectral density of PA output for matched-complexity DPD configurations: anti-diagonal constrained LS (blue) vs. $M \times K$ reduction (orange), without DPD PA output (black) for various r, M and K values denoted as r vs $M \times K$ at the top of each subplot.

TABLE I
PERFORMANCE COMPARISON OF EXISTING FPGA-BASED DPDS.

Design	r^*	Bitwidth	LUT/FF/DSP/BRAM	Power	TP	ENS/TP
Proposed	0	12-bit	1423/718/50/0	218	375	21.0
	1		1415/694/48/0	211	375	20.1
	2		1372/653/43/0	209	375	18.1
	3		1285/577/37/0	187	375	15.7
	4		1220/527/31/0	185	375	13.2
	5		1168/471/25/0	177	375	10.8
[26]	—	16-bit	1826/2815/106/0	185	400	40.9
[27]	—	—	20495/24710/504/108	—	625	154.0
[28]	—	14-bit	2298/1724/66/13	241	170	72.6
[12]*		18-bit	11254/20318/622/5†	5000	431.4	224.5
			6469/92662/796/0‡		709.7	184.6

Notes: Device platforms for [26], [27], [28], and SCALE-X are the Xilinx MPSoC Evaluation Board, Xilinx xcvu9p-flga2577-2-i, Xilinx Zynq-7Z010, and Xilinx Zynq UltraScale+ RFSoC ZU48DR, respectively. Power is in mW; throughput (TP) is in Msps. LUT/FF/DSP/BRAM denote FPGA resource counts. ENS is defined as ENS $= \max(\text{LUT}/4, \text{FF}/8) + 150 \cdot \text{DSP} + 144 \cdot \text{BRAM}$, and ENS/TP is a normalized metric. *: reports a combined system (envelope generator, leakage canceller, and DPD); DPD-only resource usage is not separately reported, †: HLS ‡: Custom RTL.

ergy savings alongside resource reduction. BRAM utilization remains at zero throughout, as coefficient storage is not the limiting factor at this model size.

B. Performance Comparison

Table I benchmarks the proposed design against four state-of-the-art FPGA DPD implementations [12], [26]–[28]. Implemented on the Zynq UltraScale+ RFSoC ZU48DR at 12-bit (Q1.11) precision, the pruned design at $r^* = 5$ uses 25 DSPs and no BRAMs, draws 177 mW at 375 Msps, and achieves the lowest ENS/TP (10.8).

The parallel architecture in [26] runs at 400 Msps but needs 106 DSPs (ENS/TP of 40.9). The sparsity-driven design in [28] has a moderate footprint, but its 170 Msps throughput yields 72.6. The design in [27] has the largest DPD-only footprint (504 DSPs, 108 BRAMs), giving 154.0. The work in [12] reports a combined envelope-generation, leakage-cancellation, and DPD system without separate DPD-only usage; it requires 622–796 DSPs depending on the design flow (HLS vs. custom RTL), with ENS/TP of 224.5 and 184.6, respectively.

Relative to [26], the proposed design uses about $4\times$ fewer DSPs at comparable throughput, with $3.8\times$ lower ENS/TP and lower power (177 vs. 185 mW). It reduces ENS/TP by $6.7\times$ and $14.3\times$ relative to [28] and [27], respectively, although [27] runs at a higher throughput. Against [12], the comparison should be read with the scope difference in mind, since that work includes a full multi-stage system. Overall, the proposed method offers the most favorable balance of resource usage, power, and throughput for resource-constrained DPD deployment.

IV. CONCLUSION

SCALE-X presented a systematic, complexity-aware, low-precision, model-based DPD framework that combined coefficient pruning with optimized FxP modeling to capture quantization effects and guide efficient design. By jointly optimizing model order, memory depth, and coefficient bitwidth, it significantly reduced computational complexity while preserving linearization performance. FPGA results confirmed reduced hardware resource usage with improved hardware efficiency.

REFERENCES

[1] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*, 2nd ed. Norwood, MA, USA: Artech House, 2006.

[2] Z. Zhu, X. Huang, M. Caron, and H. Leung, "A blind am/pm estimation method for power amplifier linearization," *IEEE Signal Processing Letters*, vol. 20, no. 11, pp. 1042–1045, 2013.

[3] M. F. Haider, F. You, S. He, T. Rahkonen, and J. P. Aikio, "Predistortion-based linearization for 5g and beyond millimeter-wave transceiver systems: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 4, pp. 2029–2072, 2022.

[4] M. T. Khan, Y. Ding, and G. Goussetis, "Next-gen digital predistortion from hardware acceleration of neural networks: Trends, challenges, and future," *IEEE Transactions on Neural Networks and Learning Systems*, 2026.

[5] H. Huang, J. Xia, and S. Boumaiza, "Novel parallel-processing-based hardware implementation of baseband digital predistorters for linearizing wideband 5g transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 9, pp. 4066–4076, 2020.

[6] W. Li, E. Guillena, G. Montoro, and P. L. Gilabert, "Fpga implementation of memory-based digital predistorters with high-level synthesis," in *2021 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)*. IEEE, 2021, pp. 37–40.

[7] B. Chen, M. T. Khan, G. Goussetis, M. Sellathurai, Y. Ding, J. F. Mota, and J. Lee, "Comet: Co-optimization of cnn models using efficient-hardware obc techniques," *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2026.

[8] M. A. Alhartomi, M. T. Khan, S. Alzahrani, A. Alzahmi, R. A. Shaik, J. Hazarika, R. Alsulami, A. Alotaibi, and M. Al-Harthi, "Low-area and low-power vlsi architectures for long short-term memory networks," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 13, no. 4, pp. 1000–1014, 2023.

[9] D. R. Morgan, Z. Ma, J. Kim, M. G. 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, 2006.

[10] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic deviation reduction-based volterra behavioral modeling of rf power amplifiers," *IEEE Transactions on microwave theory and techniques*, vol. 54, no. 12, pp. 4323–4332, 2006.

[11] T. Cappello, G. Jindal, J. Nunez-Yanez, and K. Morris, "Power consumption and linearization performance of a bit-and frequency-scalable am/am am/pm pre-distortion on fpga," in *2022 International Workshop on Integrated Nonlinear Microwave and Millimetre-Wave Circuits (INMMiC)*. IEEE, 2022, pp. 1–3.

[12] W. Li, N. Bartzoudis, J. R. Fernández, D. López-Bueno, G. Montoro, and P. L. Gilabert, "Fpga implementation of a linearization system for wideband envelope tracking power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 71, no. 4, pp. 1781–1792, 2022.

[13] R. Braithwaite, "Digital predistortion of an rf power amplifier using a reduced volterra series model with a memory polynomial estimator," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 10, pp. 3613–3623, 2017.

[14] C.-F. Cheang, P.-I. Mak, and R. P. Martins, "A hardware-efficient feedback polynomial topology for dpd linearization of power amplifiers: Theory and fpga validation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 9, pp. 2889–2902, 2018.

[15] Q. A. Pham, D. Lopez-Bueno, T. Wang, G. Montoro, and P. L. Gilabert, "Partial least squares identification of multi look-up table digital predistorters for concurrent dual-band envelope tracking power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 12, pp. 5143–5150, 2018.

[16] L. Guan, C. Yu, and A. Zhu, "Bandwidth-constrained least squares-based model extraction for band-limited digital predistortion of rf power amplifiers," in *2012 Workshop on Integrated Nonlinear Microwave and Millimetre-wave Circuits*. IEEE, 2012, pp. 1–3.

[17] M. T. Khan and O. Gustafsson, "Analyzing step-size approximation for fixed-point implementation of lms and blms algorithms," in *2023 IEEE Nordic Circuits and Systems Conference (NorCAS)*. IEEE, 2023, pp. 1–5.

[18] H. Enzinger, K. Freiberger, and C. Vogel, "Competitive linearity for envelope tracking: Dual-band crest factor reduction and 2d-vector-switched digital predistortion," *IEEE Microwave Magazine*, vol. 19, no. 1, pp. 69–77, 2017.

[19] W. Chen, S. Zhang, Y.-J. Liu, F. M. Ghannouchi, Z. Feng, and Y. Liu, "Efficient pruning technique of memory polynomial models suitable for pa behavioral modeling and digital predistortion," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 10, pp. 2290–2299, 2014.

[20] J. A. Becerra, M. J. Madero-Ayora, and C. Crespo-Cadenas, "Comparative analysis of greedy pursuits for the order reduction of wideband digital predistorters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 9, pp. 3575–3585, 2019.

[21] F. M. Barradas, T. R. Cunha, P. M. Lavrador, and J. C. Pedro, "Polynomials and luts in pa behavioral modeling: A fair theoretical comparison," *IEEE Transactions on microwave theory and techniques*, vol. 62, no. 12, pp. 3274–3285, 2014.

[22] A. Zhu, "Decomposed vector rotation-based behavioral modeling for digital predistortion of rf power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 2, pp. 737–744, 2015.

[23] Y. Li, X. Wang, and A. Zhu, "Reducing power consumption of dpd for rf pas using real-time model switching," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 3, pp. 1500–1512, 2021.

[24] M. T. Khan and J. Hazarika, "An area and energy efficient serial-multiplier," *IEEE Embedded Systems Letters*, vol. 16, no. 4, pp. 425–428, 2024.

[25] M. Rathod, A. Hazarika, A. Agrawal, N. Kumar *et al.*, "Power efficient multiplier design for error resilient edge applications," *IEEE Embedded Systems Letters*, vol. 18, no. 1, pp. 23–26, 2025.

[26] H. Huang, J. Xia, and S. Boumaiza, "Parallel-processing-based digital predistortion architecture and FPGA implementation for wide-band 5G transmitters," in *2019 IEEE MTT-S International Microwave Conference on Hardware and Systems for 5G and Beyond (IMC-5G)*. IEEE, 2019, pp. 1–3.

[27] M. Jiang, M. Li, Z. Zhao, H. Liang, X. Luo, S. Kong, and T. Cai, "A low-complexity digital predistortion technology and FPGA implementation for wide-band satellite communication," *International Journal of Circuit Theory and Applications*, vol. 54, no. 1, pp. 485–497, 2026.

[28] M. Versluis, Y. Wu, and C. Gao, "SparseDPD: A sparse neural network-based digital predistortion FPGA accelerator for RF power amplifier linearization," in *2025 35th International Conference on Field-Programmable Logic and Applications (FPL)*. Leiden, The Netherlands: IEEE, 2025, pp. 154–158.