

Hardware And Latency Optimisation for 5G Digital Pre-Distortion

Declan Byrne, Ronan Farrell and John Dooley

Maynooth University

Maynooth, Ireland

declan.byrne@mu.ie, ronan.farrell@mu.ie, john.dooley@mu.ie

Abstract—In modern radio frequency (RF) transceivers the power amplifier (PA) is a central component in terms of power consumption. Achieving efficient performance in this component results in the PA output signal becoming distorted. Linearisation can be performed using techniques such as Digital Pre-Distortion (DPD). The pre-distorter operation of a DPD system involves the constant computation of a distorted signal to ensure linear operation of the nonlinear power amplifier. In this work a novel polynomial evaluation scheme is proposed to optimise the pre-distorter operation within a DPD system. Improvements to latency and hardware requirements are possible with new techniques. Validation of the proposed design was conducted using FPGA implementations and compared to incumbent pipelined solutions for both low latency and hardware efficiency. The proposed method indicated hardware savings of 67.8%, while operating 58.7% faster, compared to an existing implementation.

Index Terms—digital pre-distortion, polynomial evaluation, 5G, FPGA, DSP.

I. INTRODUCTION

Modern cellular networks employ PAs to amplify communication signals before transmission. In order to save power and reduce heat dissipation, PAs can be driven in a region of efficient operation. However, this introduces nonlinear distortion to the PA output signal. Using linearisation techniques, distortion can be mitigated, allowing the PA to operate efficiently and linearly.

Digital Pre-Distortion (DPD) is the predominant linearisation technique used in cellular communications [1]. This method weights the PA input signal inversely to the gain of the PA. This results in the PA output signal becoming linearised.

DPD is commonly performed by employing a learning architecture, such as Fig. 1, to train a nonlinear basis function which is then applied to the PA input. Polynomial models, derived from the Volterra series, are typically chosen and referred to as DPD structures [2]. PA input and output signals are sampled and used alongside a DPD structure to extract coefficients of the polynomial function. This is performed in the DPD system element known as the post-distorter. These coefficients are then applied in the pre-distorter continuously to pre-distort the incoming PA input signal.

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) and is co-funded under the European Regional Development Fund under Grant Number 13/RC/2077.

To date, look up table (LUT) or pipelined implementations have been used [3]. Pre-distorters implemented using LUTs are suitable for low precision linearisation. Pipelined solutions meet the need for high precision linearisation, but can require extensive hardware to implement if unoptimised [4].

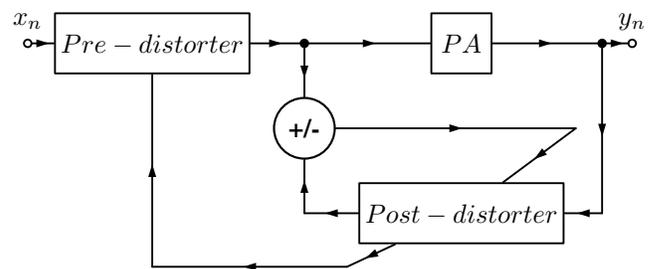


Fig. 1: Indirect learning architecture

5G cellular communications specifications have highlighted power efficient and low latency solutions as targeted goals for next generation wireless networks [5]. In a DPD system, the pre-distorter block is responsible for the majority of power consumption and latency. It is active continuously after the coefficients are trained. In contrast, the post-distorter block is only active periodically to train a set of coefficients.

Employing DPD structures using high orders of nonlinearity are necessary to linearise high efficiency PAs [6]. However, unoptimised DPD solutions tackling high order distortion do not scale well in terms of power consumption [1].

As pre-distorters apply polynomial based functions, optimisation can be achieved using polynomial evaluation methods. Polynomial evaluation schemes are algorithms which allow polynomials to be expressed in a manner such that hardware requirements and/or latency are reduced [4], [7]. The optimal scheme in terms of hardware, and thereby power, efficiency, Horner's method, has been discovered and well researched [8], [9]. However, this scheme is sequential and slow, which is counterproductive to 5G goals.

Current polynomial evaluation techniques were originally developed with a focus on optimising basic univariate polynomials. Deviating from this format (including memory taps and/or multivariate coefficients, pruning coefficients) can restrict the performance of these schemes and is common among polynomial functions applied within DPD pre-distorters.

The contribution of this paper is a polynomial evaluation scheme that takes advantage of the aforementioned factors concerning DPD functions. This proposed method presents the most favourable solution to targets set forward by 5G specifications, namely combined improved hardware efficiency and latency performance. This method also scales favourably with increasing nonlinearity order, necessary to linearise certain high efficiency PAs.

The remainder of the paper is structured as follows: In Section II a selection of polynomial structures are introduced. These are the Memory Polynomial (MP) and Generalised Memory Polynomial (GMP), commonly used DPD structures due to their compact size and high accuracy. Three existing evaluation schemes and a fourth novel method are described in Section III. Results from FPGA implementations are presented in Section IV for incumbent polynomial techniques and compared to the novel method proposed in this paper. Section V concludes the paper.

II. DIGITAL PRE-DISTORTION

Polynomial based DPD is performed by applying a set of trained coefficients to a polynomial function. The output of this function is subsequently applied as the input to the PA. Beginning with the Volterra series, shown in (1).

$$y_n = \sum_{m=1}^M \sum_{i_1=0}^K \cdots \sum_{i_m=0}^K h_m(i_1, \dots, i_m) \prod_{j=1}^m x(n - i_j). \quad (1)$$

Where y is the PA output signal, x is the PA input signal and $h_m(i_1, \dots, i_m)$ is m th order Volterra kernel. The number of coefficients used in a function can be calculated as a combination of the nonlinear order, M , and the number of memory taps, K . The nonlinear order of a model mitigates intermodulation distortion. Memory taps account for memory effects which occur within the PA during operation [10]. The amount of memory taps used in a DPD structure is often referred to as its memory depth.

A more compact polynomial structure which can still account for memory effects is the MP [11]. This DPD structure achieves good accuracy using a moderate number of coefficients, compared to others. The MP model is shown in (2).

$$y_n = \sum_{m=0}^M \sum_{k=0}^K h_{mk} \cdot x(n - m) \cdot |x(n - m)|^{k-1}. \quad (2)$$

The memory depth of the DPD coefficient, h , is denoted by the value of its subscript. The number of variables in the coefficient subscript indicates its nonlinear order. These parameters describe the signal preparation necessary to apply the coefficient. For instance, h_{11} requires a second order signal with a memory depth of two to implement.

A MP model describes a PA using an univariate polynomial. A memoryless univariate polynomial with a nonlinear order of five can be seen in (3).

$$y_n = h_{00000}x_n^5 + h_{0000}x_n^4 + h_{000}x_n^3 + h_{00}x_n^2 + h_0x_n. \quad (3)$$

Polynomial (3), and all other polynomials in this paper, will be written in a concise form shown in (4).

$$y_n = h_0^5x_n^5 + h_0^4x_n^4 + h_0^3x_n^3 + h_0^2x_n^2 + h_0x_n. \quad (4)$$

It has been shown that odd order coefficients are typically more prominent, in terms of the nonlinear response that must be corrected, compared to even order terms. Disregarding even order coefficients provides a latency and hardware resource reduction with a slight loss in performance [12]. Dropping the even order coefficients of (4) leads to polynomial (5).

$$y_n = h_0^5x_n^5 + h_0^3x_n^3 + h_0x_n. \quad (5)$$

Memory taps can be added to a DPD structure to account for the memory effects which occur in PAs. This change is reflected in the pre-distorter with the introduction of additional coefficients. Adding two memory taps to the polynomial shown in (5) forms (6).

$$\begin{aligned} y_n &= h_0x_n + h_1x_{n-1} + h_2x_{n-2} \\ &+ h_0^3x_n^3 + h_1^3x_{n-1}^3 + h_2^3x_{n-2}^3 \\ &+ h_0^5x_n^5 + h_1^5x_{n-1}^5 + h_2^5x_{n-2}^5. \end{aligned} \quad (6)$$

To further increase performance cross term coefficients can also be included in the DPD structure. Structures that include cross term coefficients include the Volterra series and GMP, presented in (7).

$$\begin{aligned} y_n &= \sum_{m=0}^M \sum_{k=1}^K h_{mk} \cdot x(n - m) \cdot |x(n - m)|^{k-1} \\ &+ \sum_{m=0}^M \sum_{k=2}^K \sum_{p=1}^P h_{mkp} \cdot x(n - m) \cdot |x(n - m - p)|^{k-1} \\ &+ \sum_{m=0}^M \sum_{k=2}^K \sum_{q=1}^Q h_{mkq} \cdot x(n - m) \cdot |x(n - m + q)|^{k-1}. \end{aligned} \quad (7)$$

Including cross term coefficients in a DPD structure can considerably increase the number of DPD coefficients. As an example, the polynomial shown in (6) can be expanded to include multivariate cross terms yields the Volterra series structure as shown in (8).

$$\begin{aligned} y_n &= h_0x_n + h_1x_{n-1} + h_2x_{n-2} \\ &+ h_0^3x_n^3 + h_0^2h_1x_n^2x_{n-1} + \cdots + h_2^3x_{n-2}^3 \\ &+ h_0^5x_n^5 + h_0^4h_1x_n^4x_{n-1} + \cdots + h_2^5x_{n-2}^5. \end{aligned} \quad (8)$$

It can be seen in polynomial (8) that many coefficients require specific signals or signal combinations in order to be applied. Delayed signals, such as x_{n-1} and x_{n-2} , are obtained using complex shift registers while signal combinations are acquired using complex multipliers. Signal preparation can be conducted using either hardware efficient or low latency orientations.

Using a hardware efficient orientation, second order signals required by a MP system with a memory depth of two can be prepared as seen in Fig. 2. This orientation uses complex shift registers to minimise the use of complex multipliers.

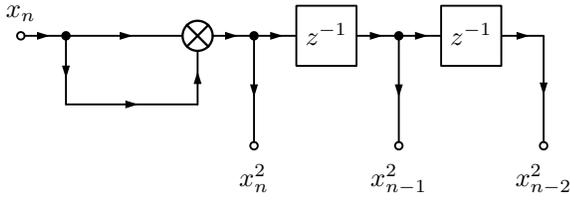


Fig. 2: 2nd order hardware efficient MP signal preparation

The signal preparation illustrated in Fig. 2 can also be accomplished using a low latency orientation as seen in Fig. 3. This orientation uses complex multipliers to prepare signals in parallel.

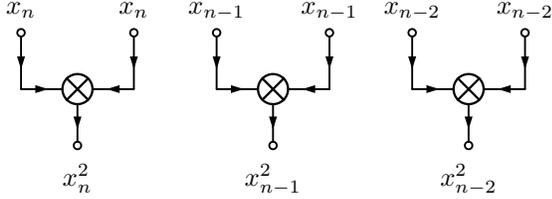


Fig. 3: 2nd order low latency MP signal preparation

Signal preparation becomes more complex for Volterra structures as the amount of coefficients grows rapidly with regard to nonlinear order. Hardware efficient preparation of second order signals in a Volterra structure with a memory depth of two can be seen in Fig. 4.

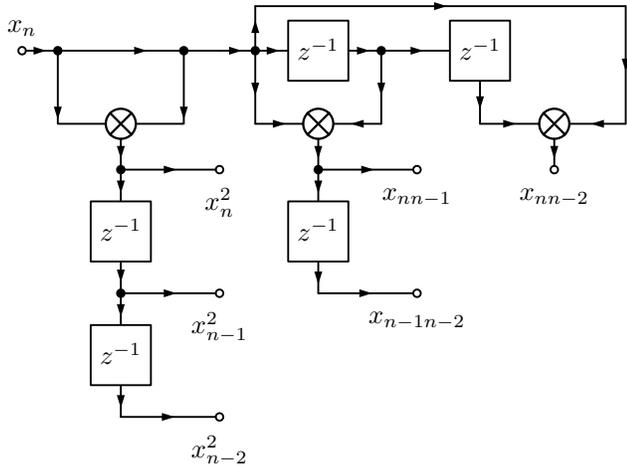


Fig. 4: 2nd order hardware efficient Volterra signal preparation

For multivariate signal preparation using the hardware efficient orientation, signals of the order M are required to obtain signals of the order $M + 1$. Each signal is multiplied with x_n and then shifted, if possible. This results in very long critical paths for DPD systems of high nonlinear order. The low latency orientation is performed identically for multivariate coefficients as it is for univariate coefficients, as seen in Fig. 3.

Polynomial evaluation schemes can simultaneously reduce computational overhead and latency costs dramatically. This

is largely accomplished through indirectly applying prepared signals to coefficients, which will be explored in the next section.

III. POLYNOMIAL EVALUATION

Existing polynomial evaluation schemes include Horner's, Dorn's and Estrin's [7]. Optimisations are made by applying low order signal terms multiple times to higher order coefficients so that signal preparation costs can be mitigated. Adders are also used in the application of coefficients instead of multipliers wherever possible. In the following subsections, the aforementioned methods and a novel proposed method are explored individually and an implemented example of the polynomial in (4) is provided for each.

A. Horner's Method

Horner's method [8] is a sequential polynomial evaluation method that obtains the optimal solution in terms of hardware requirements [9]. Due to its structure of sequential operations this evaluation scheme can be adapted for multivariable and multivariate polynomials. However, implementations using Horner's method tend to have longer critical paths compared to other methods, presenting high latency solutions. Polynomials expressed using Horner's method take the form as shown in (9) [7].

$$y = (((a_k \cdot x + a_{k-1}) \cdot x + a_{k-2}) \cdot x + \dots + a_0). \quad (9)$$

Using this scheme, polynomial (4) can be rewritten as shown in (10).

$$y_n = (((h_0^5 x_n + h_0^4) x_n + h_0^3) x_n + h_0^2) x_n + h_0 x_n. \quad (10)$$

A hardware implementation of (4) using Horner's method is shown in Fig. 5.

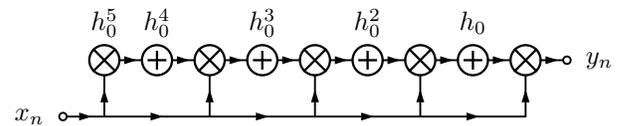


Fig. 5: Horner's method

B. Dorn's Method

Dorn's method [13] is a parallelised adaptation of Horner's method that offers improved speed performance at the cost of additional hardware resources. Polynomials expressed using Dorn's method take the form as shown in (11) [7].

$$y = q_0 + q_1 x + \dots + q_{k-1} x^{k-1}. \quad (11)$$

Where,

$$\begin{aligned} q_0 &= a_0 + a_k x^k + a_{2k} x^{2k} + \dots, \\ q_1 &= a_1 + a_{k+1} x^k + \dots, \\ &\vdots \\ q_{k-1} &= a_{k-1} + a_{k+k-1} x^k + \dots \end{aligned} \quad (12)$$

Polynomial (4) can be represented using this method as seen in (13).

$$y_n = (h_0^5 x_n^3 + h_0^2) x_n^2 + (h_0^4 x_n^3 + h_0) x_n + h_0^3 x_n^3. \quad (13)$$

A hardware implementation of (4) using Dorn's method is shown in Fig. 6.

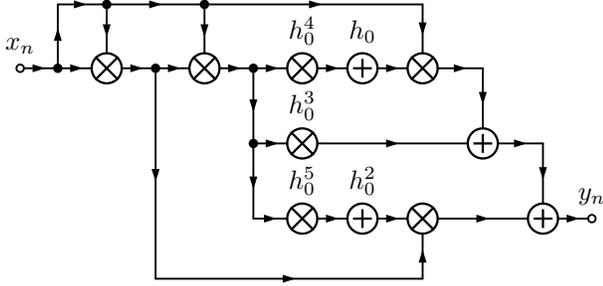


Fig. 6: Dorn's method

C. Estrin's Method

Estrin's method [14] is a parallel polynomial evaluation scheme that focuses on preparing even-order signal terms (x^2 , x^4 etc.). This process creates low latency, but hardware inefficient, implementations. Polynomials expressed using Estrin's method take the form (14) [7].

$$y = qx^{(\frac{n}{2})+1} \dots + r. \quad (14)$$

Where,

$$q = a_n x^{\frac{n}{2}} + \dots + a_{\frac{n}{2}+1}, \quad (15)$$

$$r = a_{\frac{n}{2}} x^{\frac{n}{2}} + \dots + a_0.$$

Polynomial (4) can be represented using this method as can be seen in (16).

$$y_n = (h_0^5 x_n + h_0^4) x_n^4 + (h_0^3 x_n + h_0^2) x_n^2 + h_0 x_n. \quad (16)$$

A hardware implementation of (4) using Estrin's method is shown in Fig. 7.

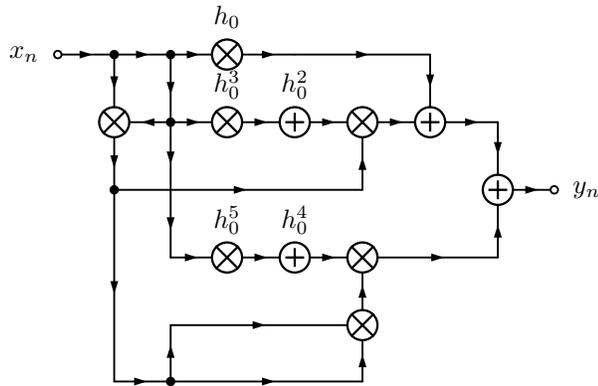


Fig. 7: Estrin's method

D. Proposed Method

The proposed method, unlike the alternatives, factorises polynomials such that even and odd order coefficients are isolated from each other along parallel paths. In the event that even order coefficients are pruned from a polynomial, entire paths are removed from the implementation. As this is a common operation in DPD functions, the proposed method presents a combined hardware efficient and low latency solution. This Polynomials expressed using the proposed method take the form (17).

$$y = (a_0 + a_1 x) + (a_4 x^2 + a_2) x^2 + (a_5 x^2 + a_3) x^3 + \dots + (a_{k+2} x^2 + a_k) x^k. \quad (17)$$

Using this method polynomial (4) can be represented as (18).

$$y_n = (h_0^5 x_n^2 + h_0^3) x_n^3 + (h_0^4 x_n^2 + h_0^2) x_n^2 + h_0 x_n. \quad (18)$$

A hardware implementation of (4) using the proposed method is shown in Fig. 8.

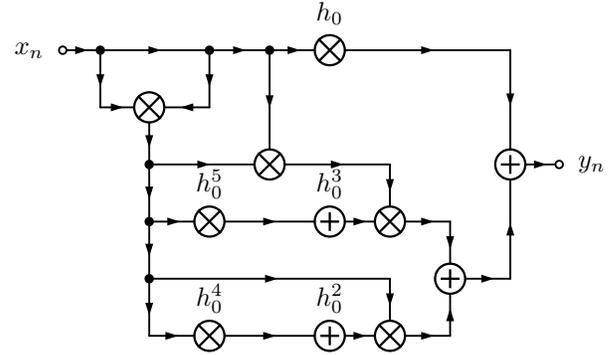


Fig. 8: Proposed method

IV. FPGA IMPLEMENTATIONS

As optimisations of polynomial functions, the discussed polynomial evaluation methods can be easily adapted to FPGA implementations, and have been in previous publications [15]. The following graphs, Figs. 9, 10, 12 and 13, describe the hardware requirements and critical path lengths of several implementations of polynomials using the described evaluation methods. These pipelined implementations were comprised of sixteen bit complex multipliers, adders and shift registers. Pipelining is performed to increase throughput, allowing multiple instructions to be performed per each clock cycle, and implemented by inserting pipeline registers in each distinct sub-stage of the datapaths [16].

Hardware requirements were measured in the amount of DSP48E1 slices used. Latency was gauged as critical path lengths and measured in terms of the number of pipeline registers contained on the path. The DSP48E1 is a type of DSP slice, programmable arithmetic units found in Xilinx FPGAs that enable fast digital signal processing [17].

A. Univariate Implementations

Univariate polynomials are applied to the PA input signal in a DPD system using the MP structure, among others. Signal preparation was conducted using a low latency approach, similar to Fig. 3. Fig. 9 shows the DSP slices required for implementing univariate polynomials of differing odd orders, such as (6), each with a memory depth of two.

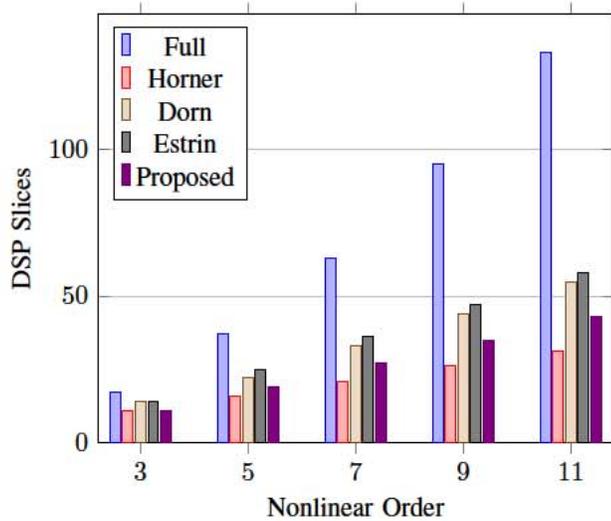


Fig. 9: Univariate implementations hardware requirements

As the nonlinear order increases the proposed method scales favourably in terms of hardware requirements, significantly better than Dorn's and Estrin's. Horner's scheme is the best performing method. However its sequential design leads to poor latency performance, as shown in Fig. 10. This graph shows the critical path lengths of the implementations described in Fig. 9.

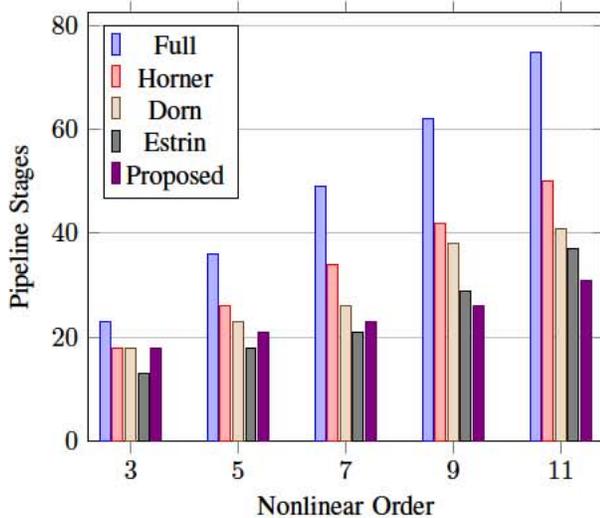


Fig. 10: Univariate implementations critical path lengths

Although the optimal method in terms of hardware efficiency, . The proposed method is the fastest method for high orders of nonlinearity. It is 19% faster than Estrin's, 32% faster than Dorn's and 61% faster than Horner's. Coupled with its competitive hardware efficiency, second only to Horner's, the proposed scheme can be identified as the favourable compromise for the design of hardware efficient, low latency DPD pre-distorters.

B. Multivariate Implementations

Multivariate polynomials are applied to the PA input signal in a DPD system using the Volterra series and certain subset structures, such as the GMP. Implementing multivariate polynomials, such as (8), in a pipelined solution expands a single complex component of a univariate implementation into arrays. Common roots across multivariate signal terms allow the polynomial evaluation schemes to be used efficiently. These roots can be determined using a multivariate oriented tree, such as Fig. 11.

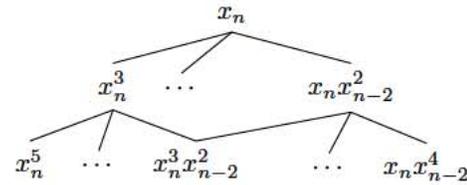


Fig. 11: Multivariate oriented tree

Signal preparation was conducted using a low latency approach, similar to Fig. 3. Fig. 12 shows the hardware requirements for implementations of odd order multivariate Volterra polynomials, such as polynomial (8), each with a memory depth of two.

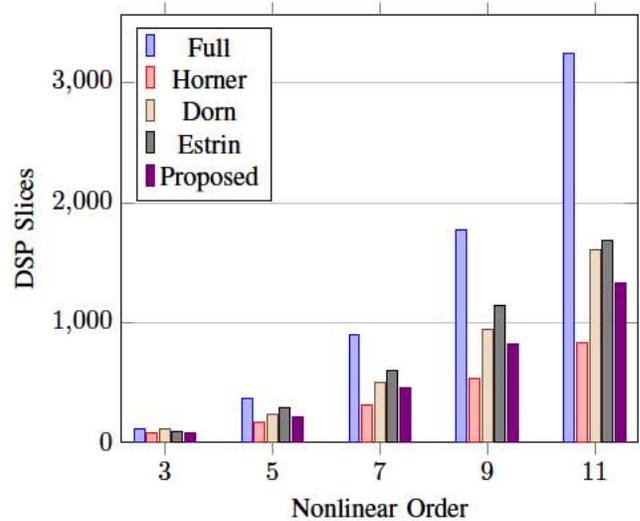


Fig. 12: Multivariate implementations hardware requirements

Applying multivariate polynomials is much more complex than in the univariate case, reflected in the general increase of hardware requirements for implementations in Fig. 12. The proposed method maintains better hardware efficiency, consistently across the range of nonlinear orders, compared to Dorn's and Estrin's schemes. Again however, Horner's method is the most hardware efficient choice, although it suffers from severe latency issues as shown in Fig. 13. This graph illustrates the critical path lengths of the implementations described in Fig. 12.

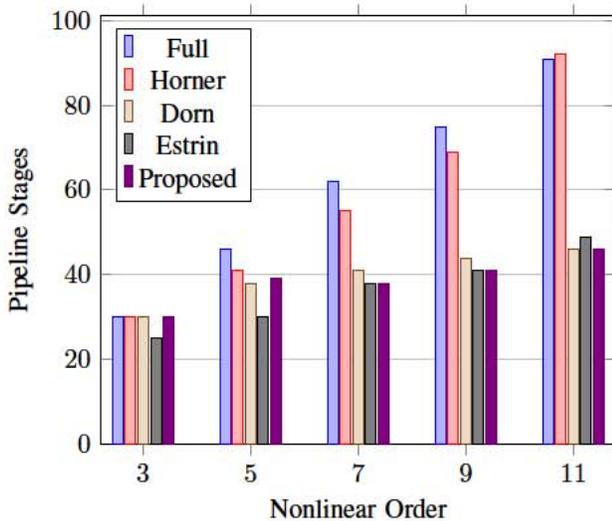


Fig. 13: Multivariate implementations critical path lengths

As shown in Fig. 13, the proposed method maintains low latency performance as the nonlinear order increases. Its performance is comparable to Dorn's and Estrin's method in terms of latency, but it is up to twice as fast as Horner's method. Due to its competitive performance in the implementation of high order polynomials, in terms of both hardware efficiency (second most efficient method) and low latency performance (joint fastest method), the proposed method is a favourable compromise for the design of hardware efficient, low latency DPD pre-distorters.

V. CONCLUSION

This work proposes a novel polynomial evaluation scheme to develop a combined hardware efficient, low latency DPD pre-distorter in accordance with 5G communication goals. This proposed method was compared to the existing polynomial evaluation techniques in terms of latency and required hardware resources. Although second in terms of hardware efficiency to Horner's scheme, the proposed method remains competitive in this category without sacrificing latency performance. This is especially true in the multivariate scenario, where the proposed scheme is up to twice as fast as Horner's method. Compared to the other methods, Dorn's and Estrin's, the proposed method remains competitive at low orders of nonlinearity, but largely outclasses these methods at

high orders of nonlinearity. High orders of nonlinearity must be accounted for in DPD structures for accurate linearisation of certain high efficiency PAs.

Overall the proposed method is the most favourable solution to the combined hardware efficiency, low latency target set forward by 5G specifications. This combination is also particularly sought in high throughput applications, that also desire low package size and power consumption, such as high throughput satellites [18].

REFERENCES

- [1] J. Wood, "Digital pre-distortion of rf power amplifiers," in *2017 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)*, Jan 2017, pp. 1–3.
- [2] K. Finnerty, J. Dooley, R. Wesson, M. P. van der Heijden, M. Acar, and R. Farrell, "Behavioral modeling of outphasing amplification systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 12, pp. 4165–4173, Dec 2016.
- [3] J. Wood, "System-level design considerations for digital pre-distortion of wireless base station transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 5, pp. 1880–1890, May 2017.
- [4] S. Xu, "Efficient polynomial evaluation algorithm and implementation on fpga," Master's thesis, Nanyang Technological University, 2013.
- [5] Nokia, "5g for mission critical communication achieve white paper," Tech. Rep., 2016.
- [6] J. Kim, B. Fehri, S. Boumaiza, and J. Wood, "Power efficiency and linearity enhancement using optimized asymmetrical doherty power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, no. 2, pp. 425–434, Feb 2011.
- [7] I. Munro and M. Paterson, "Optimal algorithms for parallel polynomial evaluation," *Journal of Computer and System Sciences*, vol. 7, no. 2, pp. 189 – 198, 1973.
- [8] W. G. Horner, "A new method of solving numerical equations of all orders, by continuous approximation," *Philosophical Transactions of the Royal Society of London*, 1819.
- [9] A. Borodin, "Horner's rule is uniquely optimal," in *Theory of Machines and Computations*, Z. Kohavi and A. Paz, Eds. Academic Press, 1971, pp. 45 – 58.
- [10] S. Boumaiza and F. M. Ghannouchi, "Thermal memory effects modeling and compensation in rf power amplifiers and predistortion linearizers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 12, pp. 2427–2433, Dec 2003.
- [11] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electronics Letters*, vol. 37, no. 23, pp. 1417–1418, Nov 2001.
- [12] N. Lashkarian and C. Dick, "Fpga implementation of digital predistortion linearizers for wideband power amplifiers," in *Proceeding of the SDR 04 Technical Conference and Product Exposition*, 2004.
- [13] W. S. Dorn, "Generalizations of horner's rule for polynomial evaluation," *IBM Journal*, 1962.
- [14] G. Estrin, "Organization of computer systems - the fixed plus variable structure computer," 1960.
- [15] M. Abbas and O. Gustafsson, "Computational and implementation complexity of polynomial evaluation schemes," in *2011 NORCHIP*, Nov 2011, pp. 1–6.
- [16] C. V. Ramamoorthy and H. F. Li, "Pipeline architecture," *ACM Comput. Surv.*, vol. 9, no. 1, pp. 61–102, Mar. 1977. [Online]. Available: <http://doi.acm.org/10.1145/356683.356687>
- [17] Xilinx, "7 series dsp48e1 slice user guide ug479 (v1.10)," Tech. Rep., 2018.
- [18] O. B. Usman, T. Delamotte, and A. Knopp, "On the complexity of sample vs. block-based predistortion for high throughput satellites," in *2018 IEEE International Conference on Communications (ICC)*, May 2018, pp. 1–6.