FPGA Architecture for the Implementation Of Polynomial Matrix Multiplication

R. Dhana Priya, D. Devi*

Department of Electronical and Communication Engineering, Sri Krishna College Of Engineering and Technology, Coimbatore, Tamilnadu, India.

*Corresponding Author: R. Dhana Priya
E-mail: priyar473@gmail.com

Received: 14/11/2015, Revised: 13/12/2015 and Accepted: 08/03/2016

Abstract

In this paper, the Polynomial Matrix Multiplication (PMM) of polynomial vectors and/or polynomial matrices have been introduced. This method provides an improvement of the fast convolution technique to multiple inputs multiple output systems (MIMO). It is devoted to the hardware implementation of PMM. The hardware implementation of this method is achieved via partly systolic, field programmable gate array (FPGA) with highly pipelined architecture. The architecture, which is flexible in terms of the order of the input polynomial matrices, Xilinx system generator tool has been used for designing. The application to sensor array signal processing strong decorrelation is highlighted. The results are demonstrated to verify the efficient performance of the architecture. The result proved that the proposed solution gives low execution times and the number FPGA resource is less.

Keywords: Field-programmable gate array (FPGA), SBR2P, polynomial matrix multiplication (PMM), polynomial matrix computations, Xilinx system generator tool. *Reviewed by ICSET’16 organizing committee

1. Introduction

The technique of polynomial matrix is equivalent to the Eigen Value Decomposition (EVD) or Singular Value Decomposition (SVD) [1] for a scalar matrix provides growing interest in recent years. They are applied in broadband extensions of narrowband issues, which usually have been mentioned by the EVD. Applications includes Broad band sensor array signal processing (SASP) [2], [6], biomedical engineering [7], sub-band coding [13], [17], and MIMO communications and coding [8], [12]. The EVD of a Para-Hermitian system, or (PEVD) polynomial EVD matrix, produces a factorization of a Para-Hermitian polynomial matrix into a product consisting of a diagonal polynomial matrix that is pre and post multiplied by (PU) par unitary polynomial matrices. A PU polynomial matrix preserves the signal power totally at every frequency [18], and so can be viewed as a lossless filter bank (stable, all-
pass). McWhirter et al. [4] proposed a clear view of the EVD to Para-Hermitian polynomial matrices, called the (SBR2) second-order sequential best rotation algorithm. It was developed for the purpose of generating a FIR PU matrix to diagonalise the Para-Hermitian polynomial matrix [4] for the signals received by a broadband sensor array. Signals that give rise to a diagonalised Para-Hermitian matrix have the strong decorrelation property [4], [13], a desired quality as in sub-band coding [17] and in SASP [4], [5]. Due to the complexity involved in manipulating polynomial matrices, the polynomial matrix diagonalization is not suited for high-speed or real-time applications because of the computational complexity present in manipulating polynomial matrices. In [19], we described a method for parallelizing SBR2, namely SBR2P, which produced a diagonalised Para-Hermitian polynomial matrix and the corresponding FIR PU filter bank. We described the implementation in hardware with highly pipelined non systolic field-programmable gate array architecture. FPGAs provide speedup performance when compared with PC-based software. They combine the main advantages of application specific integrated circuits and digital signal processors with the potential for dynamic reconfiguration. McWhirter et al. [4] and Redifet al. [5] uses the SBR2 produced FIR PU filter bank to implement strong decorrelation on the sensor signals for the broad-band SASP problems. This operation can be viewed as the polynomial matrix multiplication of a polynomial matrix with a polynomial vector or matrix-vector PMM. We also want to apply the PU filter bank to the Para-Hermitian polynomial matrix directly, to compute PEVD factorization without having to run SBR2 each time, as in [9], [16], and [17]. These operations are realized as the product of polynomial matrices and matrix-matrix PMM. The signals can be processed in the real time systems with the help of PMM in conjunction with PEVD. Especially, PMM has been a vital processing method in exploiting the broadband signal subspace proposed by a PEVD algorithm for processing real underwater acoustics data acquired from eight element array in [5] and broadband angle of arrival proposed using four-element array [6]. It played a vital role when extracting fetal components from real time life, contaminated electrocardiogram data derived from eight electrodes in [7]. PMM provides efficient application of the PEVD identified generator and parity check polynomial matrices in [9], and for the MIMO precoders and equalizers (involving four to six channels) in [10] [12]. It has also enabled the employment of PEVD-designed for the four-channel PU filter banks for optimal sub-band Coding [15] [17]. To produce an efficient use of the SBR2P architecture in real-time applications, we require the hardware implementation of the PMM. However, no unique attention has been provided to the hardware realization of PMM. The development and application of such hardware is the aim of this paper. Specially, we consider the development of an efficient, partly systolic, and highly pipelined FPGA architecture for matrix-matrix and matrix-vector PMM operations. This architecture implementation of the fast Fourier transform (FFT) technique [20] to MIMO systems, mentioned to here as fast MIMO convolution. The architecture is used along the SBR2P architecture in [19] to obtain the PEVD factorization and PU filtering of multiple channeled data of broadband SASP. The proposed architecture provides low execution times with reduced FPGA resources. It provides good approximation to the polynomial matrix computation result provided by its double precision in MATLAB for real and complex valued data. The architecture has been designed
using the Xilinx system generator tool [21], which offers interface and a number of standard modules for high speed application. Finally, the proposed architecture is used as a generic tool for polynomial matrix operations. Therefore, it is not especially designed to use with SBR2P, nor its application limited to SASP. For example, the architecture can be used with FPGA implementation of other polynomial matrix methods, as in polynomial matrix QR decomposition algorithm in [12]. This paper provides some tutorial value, continuing in a similar way to [19]. In particular, it highlights newly developed techniques for FPGA design of numerical methods for polynomial matrix application. This paper is assembled as follows. Section II presents the literature review of PMM. Section III provides polynomial Eigen value decomposition technique. Section IV proposes FPGA architecture to realize PMM. Furthermore, in Section V the analysis of the timing performance of our architecture as well as the hardware resources used in the Xilinx system generator tool by merging Xilinx ISE DESIGN SUITE 14.5. And MAT LAB R2013a is provided. Finally, the conclusion is provided in Section VI.

2. Literature review

2.1. Polynomial Matrix Multiplication for MIMO Communications

Polynomial matrices have been used for many years in the field of control. They play a major role in the realization of multi-variable transfer functions associated with (MIMO) multiple input multiple output systems. Over the last few years they have become more widely used in the context of communications, digital signal processing. Some algorithms are specially designed for this purpose, some of the algorithms broadly used for PMM are second order best sequential rotation (SBR2) and SBR2 algorithm used in parallel (SBR2P). John G McWhirter et al [19] described a generalization of the EVD for conventional Hermitian matrices to para-Hermitian polynomial matrices. This technique is called as Polynomial Eigen Value Decomposition while the underlying algorithm is known as the SBR2 algorithm. The number of iteration required to be reduced for the respective operation, and also to prevent unwanted growth in the order of the polynomial matrix being diagonalized, the name given for the algorithm was SBR. The first stage in developing polynomial singular value decomposition (PSVD) algorithm based on the SBR concept generates an SBR algorithm for a conventional matrix. For that, it is necessary to mix the QR decomposition stage of the SVD algorithm, into Kogbetliantz process [22]. This algorithm is developed by John G McWhirter for both complex and scalar matrices. The complex case is more involved than its real counterpart because the diagonal element has to remain real throughout the process.

This algorithm avoids factorization of squaring the matrix, uses only unitary and Para unitary operations, and therefore exhibits a high degree of numerical stability. Polynomial matrix multiplication, based on the application of the fast convolution technique to MIMO systems with high accuracy. FFT is the technique used to reduce the broadband problem to narrowband from using the concept of split the data into narrower frequency bands. This will increase the hardware accuracy of the system increasing the speed of operation. The proposed architecture uses limited FPGA resources and less execution time.
This is achieved by using DSP processor in conjunction with Xilinx system generator tool. Further more accurate implementations of PMM in FPGAs are expected to increase the flexibility of the algorithm considering more no of input signals and also decreasing the dependency on degree of input polynomial.

2.2. Polynomial Matrix Multiplication for Optical Computing

Recently, there is a significant advance in optical interconnections. A fiber optic communication technology provides a combination of high bandwidth, low interference, error probability, predictable message delay, and gigabit transmission capacity. Depending on the characteristics of fiber optical communications, a number of researchers have proposed optical interconnections for the processors to connect in a parallel computer system [22]. In such a system, messages can be concurrently transmitted on an optical bus with pipelined, by consuming the advantages of predictable propagation delay, unidirectional message transmission. It is now easy to integrate both electronic data computation, optical message communications is almost parallel processing systems. But in parallel algorithms have no undefined definition, scalability and time complexity is more.

3. Polynomial Eigen Value Decomposition Technique

A polynomial matrix is a matrix with polynomial elements, or its equivalent coefficients of polynomial matrix. Denote a $p \times q$ polynomial matrix in $z^{-1}$ by

$$ A(Z) = \sum_{\tau = t_1}^{t_2} A(\tau) Z^{-\tau} \quad (I) $$

Where $\tau \in \mathbb{Z}$, $t_1 \leq \tau \leq t_2$. The polynomial matrix $A(z) \in c^{p \times q}$ has entries $a_{ik}(Z) = \sum_{\tau = t_1}^{t_2} a_{ik}(\tau) Z^{-\tau}$; transform pair in (1) is represented as $A(z) \leftrightarrow A[\tau]$. The order of the polynomial matrix $A(z)$ is $(t_2 - t_1)$, where $t_1$ and $t_2$ are not positive.
The matrices $A_{[1]} \ldots A_{[2]}$ are referred to as Coefficient matrices of $A(z)$. The number of delay units required for its realization represented as the degree of $A(z)$.

A $H(z)$ polynomial matrix $\in \mathbb{C}^{p \times p}$ is said to be PU If it satisfies $H(z)H^*(z) = \overline{H(z)H^*(z)} = I$ is the par conjugate of $H(z)$. In the Context of DSP and linear system theory, $H(z)$ represents a lossless multiple-input multiple-output transfer function [22]. A polynomial matrix $R(z) \in \mathbb{C}^{p \times p}$ is Para-Hermitian if $R(z) \overset{\text{Para}}{\rightarrow} R(z)$ for all $\tau$, $\forall \tau \in \mathbb{Z}$ and $i, k = 1, 2 \ldots, p$. An example of a Para-Hermitian polynomial matrix is the polynomial (CSD) cross-spectral density matrix.

The vector sequence $\{x[t]\}_{t=-\infty}^{\infty}$ represents the set of constant correlation matrix coefficients

$$R[\tau] = E[x[t]x^H[t-\tau]], \tau \in \mathbb{Z}$$  

Where $E[.]$ is the expectation operator.

### 3.1 Computing the PEVD

As previously mentioned, polynomial matrices are used in the application of broadband SASP. In this the signals in $x[t]$ cannot be longer to decorrelate using the EVD [4]. This problem occurs in other applications for example, the source signals are broadband which undergoes convolution mixing, and they cannot be related in terms of amplitude factors and simple phase. Instead, it is important to impose strong decorrelation [13]

$$E \left( \sum_{t} x_{[t]} x_{k[t-\tau]} \right) = 0$$  

For all $\tau$ and $i = k$. One way of assuring (4) is to find a FIR PU matrix $H[t] \in \mathbb{C}^{p \times p}$ with transform $H(z) \overset{\text{Para}}{\rightarrow} H[t]$ that diagonalizes the space–time covariance matrix $R[t] = E[x[t]x^H[t-\tau]]$. In other words, we require the transformation.

$$\sum_{t=0}^{NH} \sum_{t=0}^{NH} H[t] R[t] H^H[t-\tau]$$

$$S[t] = H[t] * R[t] * H^H[-t]$$

Where $HQ + 1$ is the number of matrix coefficients of $H[t]$, $S[t] = \text{diag} \{s_{11}[t] \ldots s_{pp}[t]\}$, $\text{diag} \{\cdot\}$ represents a diagonal matrix, and the asterisk denotes convolution.

$$S(z) = H(z)R(z)H^H(z) = \text{diag}(s_{11}(z), \ldots, s_{pp}(z))$$

This is the PEVD of $R(z)$. The Para-Hermitian polynomial matrix $S(z) \overset{\text{Para}}{\rightarrow} S[\tau]$ is the CSD matrix of the PEVD-transformed signals

$$V(t) = \sum_{t=0}^{NH} H(t)X(t-\tau) = H[t] * X[t]$$
The signals $v[t] \in C_p$, which meets strong decorrelation in (4), are the result of applying the PU matrix $H[t]$, from the transformation in (6), to the received vector $x[t]$ signals $\in C_p$. The MIMO convolution in (8) can be expressed as the product of the Para unitary matrix $H(z)$ and the power series $V(z) = \sum \{v[s]\}z^{-s}$. Thus

$$V(z) = H(z) \times (z) \quad (9)$$

3.2 Fast MIMO Convolution

The discrete-time Fourier transform (DTFT) domain represented in Equation (9) by way of the convolution theorem to MIMO systems [20]. Thus, for all normalized angular frequencies.

$$V(e^{j\Omega}) = H(e^{j\Omega}) \times (e^{j\Omega}) \quad (10)$$

Where $H(e^{j\Omega}) = H(z) \mid z = e^{j\Omega} \in C_p \times p$ are unitary matrices and $X(e^{j\Omega}) = x(z) \mid z = e^{j\Omega} \in C_p$. The above equation (10) can be uniformly sampling the DTFT, which yields

$$V(e^{j\Omega_k}) = H(e^{j\Omega_k}) \times (e^{j\Omega_k}) \quad (11)$$

For $k = 0 \ldots N-1$, where $N$ is the number of frequency bins and $\Omega_k = 2\pi k/N$. Such a transformation can be obtained by taking the $N$-point discrete

$$V[t] = IFFT_N^P(FFT_N^{PXP}[H[t]] \times FFT_N^P[x[t]]) \quad (13)$$

Where $FFT_N^P$ denotes the $N$-point Fast Fourier transform along the 3D of a $p \times p$ matrices. Similarly, $FFT_N^P$ and denoted respectively, the $N$-point FFT and inverse FFT (IFFT) along the third dimension sequence of $p$-dimensional vectors. In Fig. 2 the equation in (13) is shown in the fig. 2. In some applications, such as MIMO encoding [9], [16], we need to compute the diagonalization of Para-Hermitian polynomial matrices, as in (7), using a stored (predesigned) $H(z)$. This needs two matrix–matrix PMMs, implemented by trivial extension of (13). Consider, the transformation $C[t] = H[t] \times R[t]$, which can be represented as part of that in (6). This can be implemented as follows:

$$C[t] = IFFT_N^P(FFT_N^{PXP}[H[t]] \times FFT_N^{PXP}[R[t]]) \quad (14)$$

Where $1FFT_N^{PXP}$ denotes the $N$-point IFFT along the three dimension sequence of $p \times p$ matrices, represented in Fig. 2. The two MIMO convolution techniques described in this section are important for applying hardware implemented PEVD algorithms, especially SBR2P [19], to high-speed or real-time problems

3.3 Complexity of Fast MIMO Convolution

The calculation of the matrix–vector MIMO convolution in (8) is requiring $O(p2N^2)$ time. At the same time, matrix–matrix MIMO convolution, as in (6), requires $O(p3N^2)$ time on a sequential processor. Since, fast MIMO convolution in (13) is more efficient when $\log_2 (NH + Nx - 1) < p \min (NH, Nx - 1)$. In the computation of (8) requires only $O(p2N\log2(N))$ operations instead of $O(p2N^2)$. If it is implemented on a parallel processing systems like FPGA, fast MIMO convolution will require $O(p2N\log2(N)/\kappa)$ time, $\kappa$ be the number of processing
units. At present, the contribution for the hardware implementation of PMM is not done. In the following sections, we focus on implementation of PMM by parallelizing the fast MIMO convolutions of (13) and (14).

4. Proposed Architecture

The proposed architecture gives low execution times while using limited FPGA resources as it utilizes a parallel architecture. It yields a good approximation to the polynomial matrix computation provided by its double precision counterpart for both real and complex valued data [1]. The architecture has been designed using the Xilinx system generator for DSP tool, which offers a visual interface and a number of standard modules for speedy design.

The system has operations as follows:
1. Initially, the input polynomials equations are translated into the polynomial matrices/ vectors.
2. These matrices/ vectors are fetched into two systems i.e. to the FPGA and to the MATLAB.
3. PMM computations are implemented on both the systems and outputs are evaluated for the features like area, time, accuracy, etc.
4. These results are compared for a best solution with higher accuracy. Still now, no attention is provided to the hardware realization of PMM. The development of such hardware is advantage for the MIMO communication systems. McWhirter and Redif uses the SBR2-generated FIR PU filter bank to compute strong decorrelation on the array of sensor signals for the problem of broadband SASP. This operation can be viewed as the polynomial matrix multiplication of a polynomial matrix with a polynomial vector (or matrix–vector PMM). To use the PU filter bank to the Para-Hermitian polynomial matrix directly, PEVD factorization wanted to realize without running SBR2 each time. This operation can be known as the product of polynomial matrices (matrix-matrix PMM).
5. Result

The architecture for performing the PMM, whose block diagram is shown in Fig. 2, was designed on the Xilinx system generator tool, which is a general-purpose Development board [27]. As described before, the Xilinx system generator tool, which was then synthesized, placed, mapped and routed by the Xilinx ISE 14.5 tool. FPGA-in-the-loop hardware co-simulation method was used to verify the correct operation of the FPGA design. To show the performance of the proposed PMM algorithm implementation, we represent the results from the simulation defined in Section V-A. The hardware accuracy of the fast MIMO convolution technique are explained in Section V-B, whereas the amount of resources consumed in the FPGA chip and timing performance of the hardware design are presented in Section V-C.

5.1 Hardware Accuracy

In this we consider the validity of the proposed FPGA design for PMM in terms of its accuracy, compared with a double-precision version of time-domain MIMO convolution in MATLAB. Thus, the (MSE) mean squared error is used to analyze the difference between the result yielded by our design and obtained from MATLAB. The MSE is the best suitable method for capturing the difference of the error for the types of processes used in the experiments [28].

![Figure 3: Average CPU Time Required For Calculation Of Polynomial Matrix/Vector Product](image)

![Figure 4: Hardware Accuracy Of PMM](image)

5.2 Hardware Timing Performance And Resource Utilization

To analyze the complexity of calculating the PMMs, the ensemble was simulated on a PC with 8-GB RAM running at 64-bit operating system and 2.67 GHz Intel core i5 processor. We show the average CPU access time taken by our proposed algorithm running in MATLAB versus FFT length for matrix–matrix PMM (solid curve), and the MIMO filtering (dashed curve). Also included are the MATLAB execution times required by the equivalent time-domain MIMO convolution routines for comparison. As expected, the proposed PMM algorithm for matrix–
matrix is faster than the time domain counterpart. The result is the upward trend of the curves, for \( N = 64 \) shown in Fig. 3. However, there is, on average of six fold drop in this speedup, for \( N = 512 \). Note that, the CPU access time by our algorithm for matrix–vector PMM is greater than for time-domain MIMO convolution for large \( N \). The ensemble average of \( \eta_m \) and \( \eta_v \) as a function of FFT length is shown in Fig. 4. The reduced run-times achieved by our algorithm on PC shows its applicability to real-time problems.

5.3 Hardware Implementation of PMM

The simulation result of the implementation of PMM on the system generator tool is shown in Fig 5. Xilinx ISE DESIGN SUITE 14.5. And MAT LAB R2013a are combined together and configured for system generator tool. Table 2 shows the device utilization, and Table 1 shows the optimized power (in W) and delay (in ns). Fig.6. Represents the RTL Schematic of the design.

<table>
<thead>
<tr>
<th>OPTIMIZED</th>
<th>MINIMUM USED</th>
<th>MAXIMUM USED</th>
</tr>
</thead>
<tbody>
<tr>
<td>POWER (W)</td>
<td>0.027</td>
<td>0.086</td>
</tr>
<tr>
<td>CLOCK DELAY (ns)</td>
<td>0.0165</td>
<td>0.085</td>
</tr>
</tbody>
</table>

Table 1 Optimized Power And Delay Of Pmm
<table>
<thead>
<tr>
<th>LOGIC UTILIZATION</th>
<th>USED</th>
<th>UTILIZED</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO OF SLICE LUT’S</td>
<td>727</td>
<td>1%</td>
</tr>
<tr>
<td>NO OF USED MEMORY</td>
<td>576</td>
<td>1%</td>
</tr>
<tr>
<td>NO OF USED LOGIC</td>
<td>117</td>
<td>1%</td>
</tr>
<tr>
<td>NO. OF REGISTER SLICE</td>
<td>84</td>
<td>5%</td>
</tr>
<tr>
<td>NO OF 4 INPUT LUT’S</td>
<td>202</td>
<td>13%</td>
</tr>
<tr>
<td>NO OF USED SLICES</td>
<td>121</td>
<td>15%</td>
</tr>
<tr>
<td>NO OF BUFGMU’S</td>
<td>2</td>
<td>25%</td>
</tr>
<tr>
<td>NO OF BOUNDED IOB’S</td>
<td>76</td>
<td>61%</td>
</tr>
</tbody>
</table>

Table 2 Device Utilization Of Pmm
Fig.5. Implementation Of Pmm.

Fig.6. RTL Schematic
Table.3. Power Analysis Of Pmm

6. Conclusion

In this paper, we have demonstrated architecture for the FPGA implementation of PMM. The proposed architecture is based on the application of the fast convolution technique to MIMO systems, which exploits the FFT. The important contribution of this paper is the use of a partly systolic highly pipelined architecture for the realization of fast MIMO convolution. The proposed architecture consumes limited FPGA resources and has a little dependency on the input polynomial matrices order, which makes for a scalable design. Specifically, built-in FIFOs and the sizes of the block RAMs as well as the transform length of the FFT blocks, should be modified for the input matrix order. Further, the FFT blocks transform length is, runtime configurable. The data widths used in the current implementation are reduced by the Limited number of FPGA chip. However, for larger FPGA specialized for DSP implementations are employed, the data widths are increased to achieve higher accuracy. In this case, detailed investigations in determining the optimum data width versus accuracy in multiplication. This paper focused on the relevance of the proposed design to signal processing in the context of strong decorrelation. However, as reviewed in Section I, the advantages of our architecture, with PEVD architecture is believed to make a wider ranging application. For example, when employing subspacebased methods, [5]–[7], PMM is an important stage in the separation of the broadband subspaces. It is also used in the efficient calculation and application of subspace projections, such as parity check polynomial Matrices and generator in [9], for channel coding. All these suggest that our design can be applied to solve a large number of real time problems. At present, our design supports up to four signals for broadband SASP. However, with the goal of facilitating any number of signals using the same systolic array with $4 \times 4$ PEs, we expect that one can objected an effective and efficient algorithm, which will enable the continuous operation of the array on a different set of sub-matrices pulled from the input polynomial matrices in
each iteration, permits the architecture to perform PMM of all the size matrices in a certain number of iterations. The increase in the FPGA resource consumption will possibly degrade the achievable speed-up figures.

References