New systolic array architecture for finite field division

Atef Ibrahim¹,², Hamed Elsimary¹, and Fayez Gebali³
¹ Computer Engineering Department, Prince Sattam Bin Abdulaziz University, Alkharij 11942, Riyadh, Saudi Arabia
² Microelectronics Department, Electronics Research Institute, El-Tahrir St., Dokki 12622, Cairo, Egypt
³ Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, V8W 2Y2, Canada
a) attif.ali2002@yahoo.com

Abstract: This paper proposes a new systolic array architecture to perform division operations over $GF(2^m)$ based on the modified Stein’s algorithm. The systolic structure is extracted by applying a regular approach to the division algorithm. This approach starts by obtaining the dependency graph for the intended algorithm and assigning a time value to each node in the dependency graph using a scheduling function and ends by projecting several nodes of the dependency graph to a processing element to constitute the systolic array. The obtained design structure has the advantage of reducing the number of flip-flops required to store the intermediate variables of the algorithm and hence reduces the total gate counts to a large extent compared to the other related designs. The analytical results show that the proposed design outperforms the related designs in terms of area (at least 32% reduction in area) and speed (at least 60% reduction in the total computation time) and has the lowest AT complexity that ranges from 80% to 94%.

Keywords: finite field division, systolic arrays, parallel processing, hardware security, cryptosystems, hardware accelerators

Classification: Integrated circuits

References


1 Introduction

Polynomial division over finite fields is extensively used in several applications such as data coding, error detection, digital communication systems, cryptography, and signal processing. Among the arithmetic operations in finite fields, division has been considered as the most complicated and expensive finite-field operation. Therefore, there are many algorithms presented in the literature to perform field division with their high performance realization in both software and hardware. These algorithms can be computed based on several schemes such as Fermat’s little theorem [1, 2], Extended Euclid’s algorithm (EEA) [3, 4, 5, 6, 7], and Extended Stein’s algorithm (ESA) [8, 9, 10, 11, 12].

There are two hardware techniques used to implement the field division algorithms. The first technique is the conventional technique that is based on Lookup tables and is efficient for VLSI implementation of field division algorithms over GF($2^m$) for small field size $m$ [3, 6]. When $m$ gets larger, we cannot easily use this technique in VLSI implementations due to the increasing overhead cost of area. The second technique is the systolic array technique that is considered the efficient hardware technique used to implement division algorithms over GF($2^m$) for large values of field size $m$ [12, 13]. This is due to the distinguished features of the systolic architectures such as regularity, modularity, concurrency, and local communication between processing elements that makes them more suited to high performance VLSI system design.

In this paper, we propose new systolic array architecture to perform modular division over GF($2^m$) based on the ESA scheme [12]. The architecture consists of one-dimensional systolic array of processing elements (PEs) and has area and AT complexities of $O(m)$ and $O(m^2)$, respectively. The latency of the proposed design is $2m - 1$ that makes it outperforms the previously reported designs in terms of speed.

The paper is organized as follows: Section 2 provides a brief discussion about the adopted division algorithm over GF($2^m$) [12]. Section 3 explains the proposed divider systolic structure. Section 4 shows the complexities of the proposed design and compares it with the previously reported designs. Section 5 concludes this work.

2 Division over GF($2^m$)

Let $G(x) = \sum_{i=0}^{m} g_i x^i$ be the irreducible polynomial used to define a finite field over GF($2^m$), where $g_i \in \{0, 1\}$, for $0 < i < m$, and $g_m = g_0 = 1$. We can represent any field element $Q \in GF(2^m)$ in the polynomial form as $Q(x) = \sum_{i=0}^{m-1} q_i x^i$, where $q_i \in \{0, 1\}$, for $0 \leq i < m$. Also, we can represent the reciprocal polynomial $Q^*(x)$ of $Q(x)$ as $Q^*(x) = \sum_{i=0}^{m-1} q_i x^{m-i-1}$, where $q_i \in \{0, 1\}$, for $0 \leq i < m$.

In [12], the authors proposed the following division algorithm, Algorithm 1, to obtain $V(x) = A(x)/B(x) \mod G(x)$. In this algorithm, $\Delta$ represents the upper bounds on $\deg(R)$ and $\deg(S)$ of the intermediate polynomials $R$ and $S$, where $\deg()$ indicates the polynomial degree.
Algorithm 1 Finite field division algorithm [12].

1: Input: \( A(x), B(x), G(x), \) and \( \Delta \)
2: Output: \( V(x) = (A(x)/B(x)) \mod G(x) \)
3: Initialization: \( R^0(x) \leftarrow B(x), S^0(x) \leftarrow G(x), U^0(x) \leftarrow A(x), V^0(x) \leftarrow 0, P(x) \leftarrow G(x), \Delta^0 \leftarrow -1 \)
4: Algorithm:
5: for \( 1 \leq i \leq 2m - 1 \) do
6:   if \( r_{i-1}^0 = 1 \) then
7:     if \( \Delta^i - 1 < 0 \) then
8:       \( R^i(x) \leftarrow (R^{i-1}(x) + S^{i-1}(x))/x \)
9:       \( S^i(x) \leftarrow R^{i-1}(x) \)
10:      \( U^i(x) \leftarrow ((U^{i-1}(x) + V^{i-1}(x))/x) \mod P(x) \)
11:     \( V^i(x) \leftarrow U^{i-1}(x) \)
12:     \( \Delta^i \leftarrow -\Delta^{i-1} \)
13:   else
14:     \( R^i(x) \leftarrow (R^{i-1}(x) + S^{i-1}(x))/x \)
15:     \( S^i(x) \leftarrow S^{i-1}(x) \)
16:     \( U^i(x) \leftarrow ((U^{i-1}(x) + V^{i-1}(x))/x) \mod P(x) \)
17:     \( V^i(x) \leftarrow V^{i-1}(x) \)
18:     \( \Delta^i \leftarrow \Delta^{i-1} - 1 \)
19:   end if
20: else
21:     \( R^i(x) \leftarrow R^{i-1}(x)/x \)
22:     \( S^i(x) \leftarrow S^{i-1}(x) \)
23:     \( U^i(x) \leftarrow (U^{i-1}(x)/x) \mod P(x) \)
24:     \( V^i(x) \leftarrow V^{i-1}(x) \)
25:     \( \Delta^i \leftarrow \Delta^{i-1} - 1 \)
26: end if
27: end for

For large values of field size \( m \), the counter used to compute the variable \( \Delta \) will dominate the Critical Path Delay (CPD) and reduces the divider speed to a large extent. To reduce the effect of increasing CPD, \( \Delta \) is replaced by two variables \( h \) and \( D \) such that [12]: \( h \) denotes the sign of \( \Delta \) (\( h = 0 \) if \( \Delta > 0 \) and \( h = 1 \) if \( \Delta < 0 \)) and \( D \) is a one-hot \( m \) + 1-bit vector defined by \( D = 2|\Delta| = 2^{|\Delta|-1} \). This representation of \( \Delta \) replaces the counter operation with a shift for the value of \( D \) and a bit inversion for the sign \( h \) as shown in Algorithm 2.

In Algorithm 2, variables \( \overline{R}^i(x), \overline{S}^i(x), \overline{U}^i(x), \) and \( \overline{V}^i(x) \) represent the computed values of polynomials \( R(x), S(x), U(x), \) and \( V(x) \), respectively, after iteration \( i \). We can express the recurrence equations for updating \( \overline{R}^i(x), \overline{U}^i(x), \overline{S}^i(x), \) and \( \overline{V}^i(x) \) in bit-level form as:

\[
\begin{align*}
    r_{j-1}^i &= r_{j-1}^{i-1} + c3s_{j-1}^{i-1} \\
    u_{j-1}^i &= u_{j-1}^{i-1} + c3v_{j-1}^{i-1} + c4p_{j-1}^{i-1} \\
    s_{j}^i &= cS^i u_{j-1}^{i-1} + cS^i s_{j-1}^{i-1} \\
    v_{j}^i &= cS^i u_{j-1}^{i-1} + cS^i v_{j-1}^{i-1}
\end{align*}
\]
Algorithm 2 Modified Finite field division algorithm [12].

1: Input: $A(x)$, $B(x)$, $G(x)$, $h$, $D$
2: Output: $V(x) = (A(x)/B(x)) \bmod G(x)$
3: Initialization: $R^0(x) \leftarrow B(x)$, $S^0(x) \leftarrow G(x)$, $U^0(x) \leftarrow A(x)$, $V^0(x) \leftarrow 0$, $P(x) \leftarrow G(x)$, $h^0 \leftarrow 1$, $D^0 \leftarrow 2$
4: Algorithm:
5: for $1 \leq i \leq 2m - 1$ do
6: if $r^i_0 = 1$ then
7: if $h^i = 1$ and $d^i = 0$ then
8: $R^i(x) \leftarrow (R^{i-1}(x) + S^{i-1}(x))/x$
9: $S^i(x) \leftarrow R^{i-1}(x)$
10: $U^i(x) \leftarrow ((U^{i-1}(x) + V^{i-1}(x))/x) \bmod P(x)$
11: $V^i(x) \leftarrow U^{i-1}(x)$
12: $h^i \leftarrow 0$
13: else
14: $R^i(x) \leftarrow (R^{i-1}(x) + S^{i-1}(x))/x$
15: $S^i(x) \leftarrow S^{i-1}(x)$
16: $U^i(x) \leftarrow ((U^{i-1}(x) + V^{i-1}(x))/x) \bmod P(x)$
17: $V^i(x) \leftarrow V^{i-1}(x)$
18: if $h^i = 0$ and $d^i = 0$ then
19: $D^i \leftarrow D^{i-1}/2$
20: else
21: $D^i \leftarrow 2D^{i-1}$
22: $h^i \leftarrow 1$
23: end if
24: end if
25: else
26: $R^i(x) \leftarrow R^{i-1}(x)/x$
27: $S^i(x) \leftarrow S^{i-1}(x)$
28: $U^i(x) \leftarrow (U^{i-1}(x))/x \bmod P(x)$
29: $V^i(x) \leftarrow V^{i-1}(x)$
30: if $h^i = 0$ and $d^i = 0$ then
31: $D^i \leftarrow D^{i-1}/2$
32: else
33: $D^i \leftarrow 2D^{i-1}$
34: $h^i \leftarrow 1$
35: end if
36: end if
37: end for

where,

\[ c3^i = r^i_0 \]
\[ c4^i = u^i_0 + r^{i-1}_0 v^{i-1} \]
\[ c5^i = r^{i-1}_0 h^{i-1} d^{i-1}_0 \]
Converting variable $D^i$ to bit-level from will increase the complexity of hardware design, thus we preferred to keep it in the word-level form. It can be updated along with $h^i$ using the following recurrence equations:

$$\begin{align*}
D^i &= c6^i(D^{i-1}/2) + \overline{c6^i}(2D^{i-1}) \\
h^i &= \overline{c6^i}
\end{align*}$$

(8) (9)

where,

$$c6^i = (r_0^{j-1}h_0^j)\overline{d_0^{j-1}}$$

(10)

The subscript $j$ used in the previous equations refers to the $j^{th}$ coefficient bit of polynomials $R(x)$, $U(x)$, $S(x)$, $V(x)$, and $P(x)$, where $0 \leq j \leq m$. $u_0^{i-1}$, $r_0^{j-1}$, $v_0^{j-1}$, and $d_0^{j-1}$ indicate the Least Significant Bits (LSBs) of $U(x)$, $R(x)$, $V(x)$, and $D$, respectively, at iteration $i$.

### 3 Proposed systolic array design of the division algorithm

The proposed systolic array design is obtained by using an approach previously explained by the second author in [14]. This approach starts by obtaining the dependency graph (DG) for the intended algorithm and assigning a time value to each node in the DG using a scheduling function as explained by authors in [14, 15, 16, 17, 18, 20]. The approach ends by projecting several nodes of the DG to a processing element (PE) to constitute the systolic array [14, 17, 19, 20, 21, 22].

We can extract the DG from Equations (1)–(10) for $m = 5$ as shown in Fig. 1. The equations has two indexes $(i, j)$, thus the computation domain is the convex hull in the two-dimensional (2D) space and the circles in this domain define the operations. Inputs of the DG are $s_0^j$, $r_0^j$, $u_0^j$, $v_0^j$, $p_0^j$, $D_0^j$, and $h_0^j$. The nodes in right most column (gray nodes) compute control signals $c3^i$, $c4^i$, and $c5^i$. These signals are broadcasted to all nodes in the same row $i$. The red lines (slanted lines) denote the updated bits $u_{j-1}^i$ and $r_{j-1}^i$. $p_j^{2m-1}$ bit is assigned to all nodes in each column $j$. The resulted output bits at the top of DG are the last output bits of the variable $V$, $v_j^{2m-1}$.

The systolic array architecture can be extracted by choosing the scheduling vector $S = [1\ 0]$ and the projection vector $P = [0\ 1]^T$ for the DG. These vectors are obtained from applying the approach previously reported by authors in [14, 17, 20, 21, 22].

Fig. 2 shows the resulted systolic array after applying the suggested mapping vectors on the DG. It consists of $m+1$ PEs. PE0 produces control signals $c3^i$, $c4^i$, and $c5^i$ through each $i$ iteration. Also, it updates variables $D^i$, $h^i$, and the LSB of $V^i$, $v_i^j$. The bits $s_0^i$ and $r_0^i$ are not processed in PE0 because $s_0^i$ is always equal to ‘1’ and $r_0^i$ is shifted out [9]. PEj updates $r_{j-1}^i$, $u_{j-1}^i$, $s_j^i$, $v_j^i$ based on Equations (1)–(4). PEm is a simplified form of PEj and computes $u_{m-1}^j = c3^i v_{m-1}^j + c4^i$, $r_{m-1}^j = c3^i s_{m-1}^j$, $s_j^i = \overline{c3^i s_{m-1}^j}$. This simplification is resulted from the constant input values $p_m^{2m-1} = 1$, $r_{m-1}^j = 0$, $u_{m-1}^j = 0$ applied to PEm.

The hardware details of the PEs (PE0, PEj, and PEm) are shown in Figs. 3–5. Multiplexers (Muxes) controlled by $c2$ control signal are used to select between the input signals and the intermediate results generated through different $i$ iterations. $c2$ equal ‘1’ in the first clock cycle to pass the input signals and equal ‘0’ through the
remaining clock cycles to latch the intermediate results. Variables \( D_i \) and \( h_i \) are updated based on Equations (8)–(9) and control signal \( c6^i \). This signal controls the shifter shown in Fig. 3(a) to set the shift direction (right/left). The shifter hardware details are shown in Fig. 3(b).

The operation of the systolic array can be summarized as follows:

1. At time \( n = 1 \), the external control signal \( c2 \) is set to ‘1’ to pass the initial inputs to the D-FFs and registers located inside the PEs.
2. At time \( n > 1 \), control signals \( c3^i, c4^i, \) and \( c5^i \) are produced inside PE0 and transmitted to the remaining PEs. Also, \( c2 \) control signal is set to ‘0’ to latch the updated values of the intermediate variables.
3. At the last time step \( n = 2m - 1 \), the systolic array produces the output bits of \( V, v_0^j \) and \( v_j^j \), from PE\(_0\) and PE\(_j\), \( 1 \leq j \leq m - 1 \), respectively.

4 Performance analysis and comparison

Table I shows a comparison between the proposed systolic design and the related work [5, 10, 12] in terms of area requirements and delay. In this table, \( T_A \), \( T_X \), and \( T_{MUX} \) denote the delay of 2-input AND gate, 2-input XOR gate, and 2-input Multiplexer, respectively. The total gate count is estimated in terms of 2-input NAND gate based on the NanGate (15 nm, 0.8 V) Open Cell Library that is based on the NCSU FreePDK15 process kit. The area and delay information of the standard cells is given in Table II.

We notice from Table I that the area requirements and latency of the proposed design are lower than that of the compared systolic designs. The systolic arrays
presented in [5, 10] consist of $2m$ and $2m - 1$ PEs, respectively. Besides the large number of PEs used in each systolic array, each PE consists mainly of 2 parts: datapath part and control part. The control part is repeated in each PE and this increases the area complexity of each PE and hence increases the total area complexity of the systolic arrays. Also, more flip-flops (FFs) are added inside the PEs to reduce the critical path delay and this leads to increasing the area complexity and the latency of these designs. The systolic array presented in [12] consists of $m + 1$ PEs like the proposed systolic array, but it employed a pipeline interleaving technique to improve the utilization of the PEs. This resulted in increasing the number of FFs used in the systolic array to a large extent and hence increasing the latency of the resulted systolic array to be $5m - 2$. Due to this reason and the longer critical path delay of this design, the area and time complexities of this design exceeds the proposed one.

The total gate count (area) of the proposed design is lower than the other compared designs because of the reduced number of FFs required to store the intermediate variables. The reduced number of FFs leads to reducing the latency of the proposed design to be $2m - 1$. The lower critical path delay besides the reduced latency of the proposed design lead to reducing the time complexity of the proposed design to be less than the compared designs.

To quantify the analytical results obtained in Table I, Table III shows the estimated values of total gate counts (A), latency (L), critical path delay (CPD) or clock period, total computation time (T), and AT complexity for $m = 233$. The AT complexity is defined as the area (A) multiplied by the total computation time (T). The sum of critical path delay, setup delay ($T_{\text{setup}}$) and propagation delay ($T_{\text{Ck\rightarrow Q}}$) of the D-FF represents the clock period. From Table III, we notice that the proposed design outperforms other compared designs in terms of area (at least 32%
reduction in area) and speed (at least 60% reduction in total computation time) and
has the lowest AT complexity that ranges from 80% to 94%.

5 Conclusion and future work

The details of a new systolic array architecture of the extended stein’s division algorithm over $GF(2^m)$ are explained in this article. We used a previously reported systematic approach to extract the proposed design. The approach starts by obtaining the dependency graph for the intended algorithm and assigning a time value to each node in the DG using a scheduling function and ends by projecting several nodes of the DG to a processing element to constitute the systolic array. The obtained design structure has a reduced number of flip-flops required to store the intermediate variables of the algorithm and hence a significant reduction in the total gate counts compared to the related designs. The analytical results showed that the proposed design outperforms the previously reported related designs in terms of area and speed. As a future work, the proposed design and the other related designs will be described using VHDL programming language and will be implemented on FPGA and ASIC to obtain an accurate real implementation results. Due to the lower area and delay achieved by the proposed design, we expect that the total power consumption and energy of the proposed design will be lower than that of the previously reported related designs. This makes the proposed design more suitable for resource-constrained applications that have more restrictions on area and power.

Table I. Complexity comparison of different systolic designs over $GF(2^m)$.

<table>
<thead>
<tr>
<th>Design</th>
<th>Latency</th>
<th>Critical Path delay</th>
<th>Number of Flip-Flops</th>
<th>Total Gate Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daneshbeh [5]</td>
<td>5m-4</td>
<td>$2T_A + T_X + T_{MUX}$</td>
<td>36m</td>
<td>222m</td>
</tr>
<tr>
<td>Lin [10]</td>
<td>5m-3</td>
<td>$T_A + 2T_X$</td>
<td>34m-17</td>
<td>215.2m-73.1</td>
</tr>
<tr>
<td>Wu [12]</td>
<td>5m-2</td>
<td>$T_A + 2T_X + T_{MUX}$</td>
<td>10m+5</td>
<td>68.3m+96</td>
</tr>
<tr>
<td>Proposed</td>
<td>2m-1</td>
<td>$T_A + 2T_X$</td>
<td>5m+3</td>
<td>46.5m+25</td>
</tr>
</tbody>
</table>

Table II. Area and delay values in terms of 2-input NAND gate for standard cells.

<table>
<thead>
<tr>
<th></th>
<th>INV</th>
<th>AND/OR</th>
<th>XOR</th>
<th>MUX</th>
<th>Flip-Flops</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area (2-input NAND)</td>
<td>0.6</td>
<td>1.2</td>
<td>2.5</td>
<td>2.5</td>
<td>4.3</td>
</tr>
<tr>
<td>Delay (ps)</td>
<td>5.8</td>
<td>11.3</td>
<td>12.7</td>
<td>12.4</td>
<td>32.8*</td>
</tr>
</tbody>
</table>

(§) Delay of the flip-flop is equal to the sum of the propagation delay ($T_{CLK\rightarrow Q}$) and setup time ($T_{setup}$) of the flip-flop.
Table III. Estimated area and delay results of different systolic designs over GF($2^{233}$).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Daneshbeh [5]</td>
<td>51.7</td>
<td>1161</td>
<td>47.7</td>
<td>55.4</td>
<td>2864</td>
</tr>
<tr>
<td>Lin [10]</td>
<td>50.1</td>
<td>1162</td>
<td>36.7</td>
<td>42.6</td>
<td>2134</td>
</tr>
<tr>
<td>Wu [12]</td>
<td>16.0</td>
<td>1163</td>
<td>49.1</td>
<td>57.1</td>
<td>914</td>
</tr>
<tr>
<td>Proposed</td>
<td>10.9</td>
<td>465</td>
<td>36.7</td>
<td>17.1</td>
<td>186</td>
</tr>
</tbody>
</table>

Acknowledgements
This project was supported by the Deanship of Scientific Research at Prince Sattam Bin Abdulaziz University under the research project # 2017/01/6958.