Computation of Galois Field Expressions for Quaternary Logic Functions on GPUs

Dušan B. Gajić

Abstract: Galois field (GF) expressions are polynomials used as representations of multiple-valued logic (MVL) functions. For this purpose, MVL functions are considered as functions defined over a finite (Galois) field of order $p - GF(p)$. The problem of computing these functional expressions has an important role in areas such as digital signal processing and logic design. Time needed for computing GF-expressions increases exponentially with the number of variables in MVL functions and, as a result, it often represents a limiting factor in applications. This paper proposes a method for an accelerated computation of GF(4)-expressions for quaternary (four-valued) logic functions using graphics processing units (GPUs). The method is based on the spectral interpretation of GF-expressions, permitting the use of fast Fourier transform (FFT)-like algorithms for their computation. These algorithms are then adapted for highly parallel processing on GPUs. The performance of the proposed solutions is compared with referent C/C++ implementations of the same algorithms processed on central processing units (CPUs). Experimental results confirm that the presented approach leads to significant reduction in processing times (up to 10.86 times when compared to CPU processing). Therefore, the proposed approach widens the set of problem instances which can be efficiently handled in practice.

Keywords: Spectral techniques, Galois field expressions, Finite fields, Fast Fourier transform, FFT, Parallel programming, GPU computing.

1 Introduction

Functional expression over finite fields, often called Galois field (GF) expressions, are polynomials used as analytical representations of multiple-valued logic (MVL) functions. In this sense, MVL functions are viewed as functions defined over finite (Galois) fields $GF(p)$, where $p$ is the non-necessarily prime order of the field [13].

The presented research is based on a notion that time needed for computing coefficients in GF-expressions can be significantly shortened by adapting the fast Fourier transform (FFT)-like algorithms for their computation, developed

1Dept. of Computer Science, Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia; E-mail: dusan.b.gajic@gmail.com
through their spectral interpretation, to the architecture and programming model of modern graphics processing units (GPUs) [5, 8].

The motivation for the use of GPUs for computational purposes in the presented research is the following. The approach based on using GPUs for performing general-purpose non-graphics algorithms in a highly parallel manner, called general-purpose computing on GPUs (GPGPU) or GPU computing, has recently attracted significant interest of researchers [3, 12]. The fast-growing field of GPU computing was made possible by the evolution of GPU single-instruction, multiple-threads (SIMT) architectures [1, 12], followed by the appearance of Nvidia CUDA [9] and OpenCL [2, 11] programming frameworks, which made the immense GPU computational resources more accessible to researchers. Each step of the FFT-like algorithms involves performing the same elementary butterfly operations over different subsets of operands [4]. These butterfly operations are independent of each other and, therefore, can be performed in parallel, without any need for communication and synchronization within a step. These properties of the FFT-like algorithms make them a suitable match to the nature of the GPU hardware.

The remainder of the paper is organized as follows. In Section 2, we discuss the theoretical background of the paper. The proposed mappings of the Cooley-Tukey and the constant geometry fast algorithms for computing Galois field expressions for quaternary logic functions to the architecture and programming model of GPUs are considered in Section 3. Their implementation for efficient processing on graphics processors is discussed in Section 4. Section 5 presents the experimental environment, used for the verification of the proposed solutions, and the recorded experimental results. The paper is closed with some conclusions, as well as a short discussion of possible directions for further work.

2 Galois Field Expressions for Quaternary Logic Functions

In this section, we present the definition of the Galois field expressions for quaternary (four-valued) logic functions and the Cooley-Tukey and the constant geometry FFT-like algorithms for computing coefficients in these expressions. For more detailed discussions of these topics, we refer to [13, 15].

Each quaternary function of \( n \) variables can be represented as a polynomial of the form

$$f(x_1, x_2, \ldots, x_n) = \sum_{i=0}^{4^n-1} g_i \phi_i,$$

where \( g_i \in \{0, 1, 2, 3\} \), and \( \phi_i \) are the product terms defined in the natural (Hadamard) order as elements of the vector \( X_{4GF}(n) \) defined as
Computation of the Galois Field Expressions for Quaternary Logic Functions on GPUs

\[ X_{4GF} (n) = \bigotimes_{i=1}^{n} X_{4GF} (1), \quad X_{4GF} (1) = [x_i^0, x_i^1, x_i^2, x_i^3], \]  

(2)

where \( \bigotimes \) denotes the Kronecker product. Since 4 is a non-prime number, additions and multiplications in \( GF(4) \) are carried out as defined in Table 1. Notice that, when \( p \) is a prime number, ordinary operations of addition and multiplication modulo \( p \) can be used as operations in \( GF(p) \) [13].

**Table 1**

*Addition and multiplication in \( GF(4) \).*

<table>
<thead>
<tr>
<th>+</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \cdot )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

The set of basic functions for \( n = 1 \) is given by the columns of the matrix

\[ X_{4GF} (1) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 1 \\ 1 & 3 & 2 & 1 \end{bmatrix}. \]  

(3)

In matrix notation, the coefficients \( g_i \) in the Galois field expression for a function \( f \), specified by the function vector \( F = [f(0), f(1), \ldots, f(4^n - 1)]^T \), are computed as

\[ S_{f,AGF} = G_{4GF} (n) F, \]  

(4)

where

\[ G_{4GF} (n) = \bigotimes_{i=1}^{n} G_{4GF} (1), \quad G_{4GF} (1) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 3 & 2 \\ 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix}. \]  

(5)

**Example 1.** For a quaternary logic function with two variables \( f(x_1, x_2) \), given by the function vector \( F = [3, 1, 2, 0, 2, 1, 2, 2, 0, 3, 1, 0, 0, 2, 3, 2]^T \), the GF transform matrix in \( GF(4) \) can be computed directly from the definition by using (5). Then, using this matrix and (4), the coefficients in the \( GF(4) \)-expression for \( f \) are computed by matrix-vector multiplication which yields

\[ S_{f,AGF} = G_{4GF} (2) F = [3, 0, 2, 0, 2, 2, 3, 3, 2, 1, 0, 2, 1, 0, 0, 2, 2]^T. \]  

(6)

99
However, this approach has a computational complexity of $O(N^2)$, where $N = 4^n$ is the length of the function vector.

The same computations can be performed using the FFT-like algorithms, with an $O(N \log N)$ asymptotical time complexity [13, 15]. In this paper, we discuss the Cooley-Tukey and the constant geometry fast algorithms, based on different factorizations of the transform matrix [10, 15].

In the case considered in this example, the Cooley-Tukey algorithm stems from the factorization of the GF(4) transform matrix taking the following form

$$G_{4GF}(2) = C_1 C_2.$$  \hfill (7)

where

$$C_1 = G_{4GF}(1) \otimes I_4(1), \quad C_2 = I_4(1) \otimes G_{4GF}(1).$$  \hfill (8)

$I_4(1)$ in (8) represents the $4 \times 4$ identity matrix. Each of the matrices $C_1$ and $C_2$ describes a step in the Cooley-Tukey FFT-like algorithm for computing the coefficients in $S_{f,4GF}$. The basic transform matrix $G_{4GF}(1)$ specifies the elementary butterfly operation in the algorithm.

The computation through this algorithm is performed as

$$S_{f,4GF} = G_{4GF}(2)F = C_1(C_2F).$$  \hfill (9)

The first step of the algorithm computes

$$S_{f_1,4GF} = C_2F = [3, 1, 2, 0, 2, 0, 0, 1, 2, 1, 2, 3, 1, 1, 2, 0, \ldots]^T,$$  \hfill (10)

while the second step produces

$$S_{f_2,4GF} = C_1S_{f_1,4GF} = [3, 0, 2, 0, 2, 2, 3, 3, 2, 1, 0, 2, 1, 0, 2, 2, \ldots]^T.$$  \hfill (11)

The same coefficients can be computed using the constant geometry fast algorithm, based on the factorization of the transform matrix into identical factor matrices in the following way

$$G_{4GF}(2) = A^2,$$  \hfill (12)

where

$$A = \begin{bmatrix} A_0 \\ A_1 \\ A_2 \\ A_3 \end{bmatrix}, \quad A_i = \text{diag}(a_i, a_i, \ldots, a_i), \quad i = 0, 1, 2, 3,$$  \hfill (13)

where $a_i$ are the row-vectors of $G_{4GF}(2)$

$$a_0 = [1, 0, 0, 0], \quad a_1 = [0, 1, 3, 2], \quad a_2 = [0, 1, 2, 3], \quad a_3 = [1, 1, 1, 1].$$  \hfill (14)

Therefore, the factor matrix $A$, in the case of factorizing $G_{4GF}(2)$, has the form
Computation of the Galois Field Expressions for Quaternary Logic Functions on GPUs

\[
A = \begin{bmatrix}
A_0 \\
A_1 \\
A_2 \\
A_3
\end{bmatrix} = \begin{bmatrix}
a_0 & 0 & 0 & 0 \\
0 & a_0 & 0 & 0 \\
0 & 0 & a_0 & 0 \\
0 & 0 & 0 & a_0 \\
a_1 & 0 & 0 & 0 \\
0 & a_1 & 0 & 0 \\
0 & 0 & a_1 & 0 \\
0 & 0 & 0 & a_1 \\
a_2 & 0 & 0 & 0 \\
0 & a_2 & 0 & 0 \\
0 & 0 & a_2 & 0 \\
0 & 0 & 0 & a_2 \\
a_3 & 0 & 0 & 0 \\
0 & a_3 & 0 & 0 \\
0 & 0 & a_3 & 0 \\
0 & 0 & 0 & a_3
\end{bmatrix}, \quad (15)
\]

where \( \theta = [0, 0, 0, 0] \).

Then,

\[
G_{4GF}(2) = A^2. \quad (16)
\]

The use of constant geometry algorithms for computing coefficients in GF-expressions was first proposed in [14].

In this case, the first step of the algorithm yields

\[
S_{f_1,AGF} = \begin{bmatrix} 3 & 2 & 0 & 0 & 0 & 0 & 3 & 0 & 3 & 2 & 3 & 1 & 2 & 0 & 3 & 2 & 3 \end{bmatrix}^T, \quad (17)
\]

while the second step produces the same final coefficients

\[
S_{f_1,2,AGF} = AS_{f_1,AGF} = \begin{bmatrix} 3 & 0 & 2 & 0 & 2 & 2 & 3 & 3 & 2 & 1 & 0 & 2 & 1 & 0 & 2 & 2 \end{bmatrix}^T. \quad (18)
\]

Fig. 1 shows the signal flow graphs for the Cooley-Tukey and the constant geometry FFT-like algorithms for computing coefficients in a GF(4)-expression for a quaternary function of two variables. In this figure, solid lines correspond to weights of 1, dashed lines carry weights 2, and dotted lines represent weights of 3. Operands on which different elementary butterflies operate are identified by different colors in the figure.
Both the Cooley-Tukey and the constant geometry algorithms are characterized by identical asymptotical time complexity and involve performing the same elementary butterfly operations, defined by the basic transform matrix. However, differences in these two factorizations of the $GF$-transform matrix lead to distinct memory access patterns, which can have significant effects on processing times needed for computing coefficients in $GF$-expressions.

3 Mapping of the Algorithms

The key issues in efficient mapping of the FFT-like algorithms to parallel computing architectures, such as GPUs, are:

1. The implementation of address arithmetic, i.e., computations needed for data fetching and storing, and
2. Organization of computations, i.e. distribution of the butterfly operations over processing elements.
In the proposed mappings, computations involved in the elementary butterflies of the considered fast algorithms are mapped to device programs, processed on the GPU, while the rest of the procedures are performed as parts of the CPU-processed host programs [6, 9]. Each thread performs a radix-4 butterfly over different input function values and the number of threads in each algorithm step is, therefore, \( N/4 \). This organization of computations, in which each butterfly is mapped to a single GPU thread, allows having a large number of threads performing the same operations on different data simultaneously, which is a good match to the nature of the GPU hardware [3].

In both CPU and GPU implementations of the Cooley-Tukey algorithm, the index \( l \) of the memory location which holds the first operand for each butterfly is computed as

\[
l = b \% d + p \cdot d \cdot (b \div d),
\]

where \( b \) is a positive integer that uniquely identifies the butterfly, \( d \) is the distance between the elements over which the butterfly performs the computations in the current step of the algorithm, \( \% \) represents the modulo operation and \( \div \) stands for integer division. The value of \( d \) for the \( k \)-th step is computed as \( d = p^{n-k} \). Notice that both the modulo operation and the integer division are not implemented directly in CPU or GPU hardware and, therefore, are performed by using routines based on simpler arithmetic operations such as addition and multiplication [2, 9].

The other operands for a butterfly are fetched from the locations with indices \( l_i \) computed as

\[
l_i = l + i \cdot d,
\]

for \( i = 1, 2, \ldots, p-1 \). Notice that memory locations of the operands on which each butterfly operates change with each algorithm step.

For the constant geometry algorithm, the index \( l \) of the memory location which holds the first operand for each butterfly is computed as

\[
l = b \cdot p,
\]

while the rest of the operands are then fetched from locations \( l_i \) computed as

\[
l_i = l + i,
\]

for \( i = 1, 2, \ldots, p-1 \). Notice that memory locations of the operands on which each butterfly operates are independent of the current algorithm step.

Since, in the case of the constant geometry algorithms, memory locations in which results of computations are stored differ from the ones from which the input data were read, these locations need to be computed in the following way

\[
o_i = b + i \cdot p^{n-1}, \quad i = 1, 2, \ldots, p-1.
\]
After comparing (19) and (20) with (21)–(23), it can be concluded that the address arithmetic in the case of constant geometry algorithm involves simpler operations, which are directly implemented in hardware and, therefore, can be performed faster. As a consequence, processing times for these algorithms are shorter than for the Cooley-Tukey algorithms, as reported in Section 5.

4 Implementation of the Algorithms

The Cooley-Tukey FFT-like algorithms allow in-place implementations using a single array in memory, since input data are fetched from the same locations where the results are stored. However, these algorithms require more complex address arithmetic than the constant geometry algorithms, since the locations of the operands change in each step of the algorithm.

Constant geometry algorithms require an out-of-place implementation with separate memory arrays for input data and output results, since operands are read from one set of locations and results are stored in another set of locations. On the other hand, these algorithms involve simple address arithmetic.

As a consequence, GPU implementations of these two algorithms require different amounts of GPU resources, in terms of registers, number of instructions, and occupancy [6, 14].

The operations of addition and multiplication in $GF(4)$, defined as in Table 1, need to be implemented on the program level, since the syntax of programming languages includes only modulo operations which can be used for computing $GF(p)$-expressions solely for $p$ prime. This is most efficiently done by using look-up tables, represented as matrices. The result of performing the operation on two operands $x$ and $y$ is the element of the matrix stored at location with indices $x$ and $y$.

5 Experimental Results

5.1 Experimental settings

The experiments reported in this section are performed on platforms specified in Table 2. Since computations are performed over function vectors, the running time of the implementations is independent of function values. Therefore, the experiments are performed using randomly generated quaternary functions with different number of variables. The presented values for the CPU and GPU computation times are average values for 10 program executions. The discussed algorithms were also implemented as sequential C/C++ programs processed on CPUs. These implementations were developed solely to measure baseline processing times needed to establish the speed-ups achieved by transferring computations to GPUs. The source code for referent C/C++ implementations is compiled for the $x64$ platform using the maximum level of
performance-oriented optimizations. Proposed GPU implementations of the considered algorithms were developed using CUDA (for Nvidia GPU on platform Intel/Nvidia) and OpenCL (for AMD GPU on platform Intel/AMD).

Table 2
System specifications for the experimental platforms.

<table>
<thead>
<tr>
<th>Platform</th>
<th>Intel/Nvidia</th>
<th>Intel/AMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU type</td>
<td>Intel Core i7-920</td>
<td>Intel Core i7-2600K</td>
</tr>
<tr>
<td>frequency</td>
<td>2.66 GHz</td>
<td>3.4 GHz</td>
</tr>
<tr>
<td>number of cores/threads</td>
<td>4/8</td>
<td>4/8</td>
</tr>
<tr>
<td>RAM</td>
<td>12 GB DDR3 2000 MHz</td>
<td>16 GB DDR3 1600 MHz</td>
</tr>
<tr>
<td>GPU Type</td>
<td>Nvidia GTX 560 Ti</td>
<td>AMD Radeon HD 6750</td>
</tr>
<tr>
<td>Number of cores</td>
<td>384</td>
<td>720</td>
</tr>
<tr>
<td>Core frequency</td>
<td>900 MHz</td>
<td>800 MHz</td>
</tr>
<tr>
<td>Processing power</td>
<td>1.26 TFLOPs</td>
<td>1 TFLOPS</td>
</tr>
<tr>
<td>Graphics RAM</td>
<td>1 GB GDDR5 4.0 GHz</td>
<td>1 GB GDDR5 1.1 GHz</td>
</tr>
<tr>
<td>Memory bandwidth</td>
<td>128 GB/s</td>
<td>73.6 GB/s</td>
</tr>
<tr>
<td>Operating system</td>
<td>Windows 7 Ultimate (64-bit)</td>
<td></td>
</tr>
<tr>
<td>IDE</td>
<td>MS Visual Studio 2012 Ultimate</td>
<td></td>
</tr>
<tr>
<td>SDK</td>
<td>Nvidia GPU Computing 5.0</td>
<td>AMD APP 2.9</td>
</tr>
<tr>
<td>GPU driver</td>
<td>Nvidia 320.57</td>
<td>AMD Catalyst 13.9</td>
</tr>
<tr>
<td>Profiling tools</td>
<td>Nvidia Parallel Nsight 3.1</td>
<td>AMD APP Profiler 2.5</td>
</tr>
</tbody>
</table>

5.2 Experimental results

The results regarding processing times for different implementations of the considered fast algorithms for computing coefficients in Galois field expressions for quaternary functions are presented in Table 3, as well as in Figs. 2 and 3.

On the Intel/Nvidia platform, the CUDA GPU implementations of the Cooley-Tukey and the constant geometry algorithms are up to 10.22 and 10.86 times, respectively, faster than their C/C++ implementations processed on CPUs. The CPU implementation of the constant geometry algorithm is up to 13% faster than the respective Cooley-Tukey FFT implementation, while gains on the GPU are up to 7.5%. On the Intel/AMD platform, the OpenCL GPU implementations outperform their referent CPU counterparts by a factor of 3.48, for the Cooley-Tukey algorithm, and 3.97, for the constant geometry algorithm. On this platform, the CPU implementation of the constant geometry algorithm is up to 12.5% faster than the respective Cooley-Tukey algorithm implementation, while on the GPU the improvement in processing time is up to 16.7%.
Fig. 2 – Computation times for different implementations on the Intel/Nvidia platform.

Fig. 3 – Computation times for different implementations on the Intel/AMD platform.
Table 3

Computation times in [ms] for different implementations of the Cooley-Tukey (CT) and the constant geometry (CG) algorithms for computing GF(4)-expressions.

<table>
<thead>
<tr>
<th>N</th>
<th>Intel/Nvidia</th>
<th>Intel/AMD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CPU/C++</td>
<td>GPU/CUDA</td>
</tr>
<tr>
<td></td>
<td>CT</td>
<td>CG</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>5.5</td>
</tr>
<tr>
<td>10</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>11</td>
<td>76</td>
<td>74</td>
</tr>
<tr>
<td>12</td>
<td>325</td>
<td>317</td>
</tr>
<tr>
<td>13</td>
<td>1409</td>
<td>1394</td>
</tr>
</tbody>
</table>

The following can be concluded after analyzing times needed for performing individual steps of the examined algorithms, shown in Table 4 for the case of the largest considered quaternary functions ($n = 13$). Processing times for the CUDA implementation of the Cooley-Tukey FFT show a significant rise in the last two steps, leading to lower overall performance. This increase in processing times can be explained by the change in memory access pattern for storing results of computations, with writes occurring at evermore closer locations which can result in memory access conflicts. The constant geometry algorithm implementation is characterized by almost constant processing times per step, which are, at start, a bit slower than times for the Cooley-Tukey algorithm, but avoid later rise and, thus, achieve better overall performance.

Table 4

Computation times [ms] for different steps of CUDA implementations of the Cooley-Tukey (CT) and the constant geometry (CG) algorithms, in the case $n = 13$.

<table>
<thead>
<tr>
<th>step</th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
<th>7.</th>
<th>8.</th>
<th>9.</th>
<th>10.</th>
<th>11.</th>
<th>12.</th>
<th>13.</th>
</tr>
</thead>
</table>

The profiling tools used for measuring and optimizing GPU program performance (see Table 2) also allow an insight into further causes of shorter processing times for implementations of the constant geometry algorithm than for the implementations of the Cooley-Tukey algorithm. The improved processing times can also be explained by lower use of resources per thread in the GPU implementations of the constant geometry algorithm. The OpenCL implementation of the Cooley-Tukey requires 27 registers and 100 instructions
per thread, while the respective constant geometry algorithm implementation needs 24 registers and only 65 instructions per thread. For CUDA implementations, the ratio of per thread register use is 23:20 in favor of the constant geometry algorithm, leading to effective multiprocessor occupancy of 66.7%, for the Cooley-Tukey, and 100% for the constant geometry algorithm implementations.

6 Conclusion

In this paper, we proposed a method for accelerated computation of $GF(4)$-expressions for quaternary logic functions, based on the application of GPUs. The method stems from the spectral interpretation of $GF(4)$-expressions, which permits the use of FFT-like algorithms, which are, in the presented research, adapted for highly parallel processing on GPUs. We consider two classes of FFT-like algorithms – the Cooley-Tukey and the constant geometry algorithms. We show that the simpler address arithmetic in the constant geometry algorithm leads to lower use of resources, and, as a consequence, better performance on both CPUs and GPUs. Experimental results confirm that the presented approach of using GPUs for computations leads to significant reduction in processing times (up to 10.86 times when compared to referent CPU implementations). In this way, the application of the proposed method widens the set of problem instances which can be efficiently handled in practice. Plans for further work include development of hybrid algorithm implementations, which would use both CPUs and GPUs for performing computations in parallel. In addition, one of the goals for future research is the construction of implementations of external memory algorithms for GPUs, which would allow processing of larger MVL functions than the ones considered in this paper.

Acknowledgement

The author would like to thank Prof. Radomir S. Stanković for his contribution to the research reported in this paper. The presented research is partly supported by the Ministry of Education and Science of the Republic of Serbia, projects ON174026 (2011-2014) and III44006 (2011-2014).

8 References


Computation of the Galois Field Expressions for Quaternary Logic Functions on GPUs