Multiplierless Unified Architecture for Mixed Radix—2/3/4 FFTs

Fahad Qureshi, Jarmo Takala
Faculty of Computing and Electrical Engineering
Tampere University of Technology
Tampere, Finland
Email: fahad@tut.fi, jarmo.takala@tut.fi

Anastasia Volkova, Thibault Hilaire
Sorbonne Universités, UPMC Univ Paris 06,
UMR 7606, LIP6, F-75005,
Paris, France
Email: anastasia.volkova@lip6.fr, thibault.hilaire@lip6.fr

Abstract—This paper presents a novel runtime-reconfigurable, mixed radix core for computation 2−, 3−, 4− point fast Fourier transforms (FFT). The proposed architecture is based on radix-3 Wingorad Fourier transform, however multiplication is performed by constant multiplication instead of general multiplier. The complexity is equal to multiplierless 3-point FFT in terms of adders/subtractors with the exception of a few additional multiplexers. The proposed architecture supports all the FFT sizes which can be factorized into 2, 3, 4 point systems. We also show that the proposed architecture has the same bound on the accuracy as the classical one.

Index Terms—Fast Fourier transform (FFT), Memory-based FFT, Mixed radix

I. INTRODUCTION

Fast Fourier transform is an integral part of orthogonal frequency division multiplexing (OFDM) system, which is used in almost every communication application, e.g. 3GPP long term evolution (LTE), wireless personal area network (WPAN), wireless metropolitan area network (WMAN), and other mobile applications. In order to design FFT processors, various algorithms and architectures have been proposed over the last four last decades. These can be divided into two types, power-of-two and non-power-of-two. Certainly, the design of non-power-of-two FFT processors are more challenging because both data management and data processing is not regular [1]–[6].

Over time, various FFT algorithms and architectures have been proposed to design an efficient FFT processor. In order to meet the requirements of applications, mainly two principal FFT architectures are used: pipelined and memory-based [7]–[10]. In general, memory-based is preferred, as it has several advantages, such as low area or hardware resources occupied by the architecture and low power consumption with complex control logic. However, there are two important challenges during its design: one is the conflict-free memory access and other is the design of core for computation of butterflies (small point FFTs or radices). The design of the core for a single radix is relatively simple and a well-known subject. However, the design of multiradix architectures still proposes some challenges. Usually, multiradix architectures are used for hardware cost reduction purposes. This is achieved by reusing the adders and multipliers [1], [3], [11]. These cores are based on mapping the all radices onto computations on the largest radix. However, the radix-2/3/4 FFTs cannot be easily mapped on the radix-4 by the usual approach. This is due to the fact that radix-2/4 FFTs do not require any multiplication, while radix-3 has one non-trivial multiplication.

In this paper, we propose to replace the general multiplier of the radix-3 FFT by a constant multiplier and then to map the radix-2 and radix-4 on the multiplierless radix-3 architecture. As a result, we propose a core that is multiplierless and can be configured by multiplexers for multiradix purposes. This core can also be used for variable length FFTs, e.g. the 3GPP LTE requires various transform sizes: 128, 256, 512, 1024, 1536, 2048 and 12−1296 respectively [12]–[15]. Our design neither requires the separate hardware for each radix, nor requires the use of costly multiplier. Furthermore its accuracy is not worse than the accuracy of the classic implementations.

The paper is organized as follows. We briefly review the background on the individual architectures of 2, 3, and 4-point FFTs in Section II. Then, we explain the main idea behind the multiplier-less multiplication in fixed-point arithmetic in Section II-C. Then, Section III presents the proposed multiplierless architecture. In section Section IV we propose a computer arithmetic based approach on the error analysis of the obtained implementation. Finally Section V provides the main conclusions of paper.

II. TECHNICAL PRE-REQUISITES

A. Radix-2/3/4 FFT

The Cooley-Tukey FFT algorithm [16] is based on the decomposition of a large FFT into two small FFTs. This decomposition is applied recursively until it reaches the small point FFTs. The small FFTs are the basic processing elements of the FFT computation, they are often referred to as butterflies or radices.

The radix-2 FFT can be expressed as:

$$
\begin{bmatrix}
X(0) \\
X(1)
\end{bmatrix} =
\begin{bmatrix}
1 & 1 \\
1 & -1
\end{bmatrix}
\cdot
\begin{bmatrix}
x(0) \\
x(1)
\end{bmatrix},
$$

(1)

where \(x(\cdot)\) and \(X(\cdot)\) represent the vectors of the input and output sequences respectively.

The radix-3 FFT is the following computation:
In this arithmetic, real numbers are represented as integers and perform computations. In digital systems, the memory registers with finite capacity are limited. Therefore, multiplication would be easy to perform. However, in modern architectures of radix-2, 3, and 4 are based on the Cooley-Tukey FFT algorithm, whereas the radix-4 SFG is based on Winograd Fourier transform algorithm (WFTA).

B. Winograd Fourier Transform Algorithm

This algorithm has the minimum number of multiplications at the expense of introducing a few extra additions [17]. Although WFTA is very efficient for small prime size FFTs, for larger sizes the number of additions becomes too high for practical implementations. In the WFTA, the FFT computation equation is written as

\[
[X(0)\ldots X(N-1)]^T = O \cdot M \cdot I \cdot [x(0)\ldots x(N-1)]^T, \tag{4}
\]

where \( I \) is a matrix corresponding to additions between inputs, \( M \) is a diagonal matrix with the multiplications, and \( O \) is a matrix corresponding to additions after the multiplications. The advantages of such architecture is that the multiplication is either real or imaginary.

Denote \( c = r + j \cdot i \) to be a complex constant only with real part \( r = \cos \alpha, i = 0 \), and \( x = x_{re} + j \cdot x_{im} \) be a complex variable. Then, the complex product \( Y = c \cdot x \) has real and imaginary parts that are computed with

\[
\begin{bmatrix}
Y_{re} \\
Y_{im}
\end{bmatrix} =
\begin{bmatrix}
\cos \alpha & 0 \\
0 & \cos \alpha
\end{bmatrix}
\begin{bmatrix}
x_{re} \\
x_{im}
\end{bmatrix}, \tag{5}
\]

where imaginary part is zero. Conversely, multiplication of \( x \) by a constant \( c \) with only imaginary part \( r = 0, i = \sin \alpha \) is described as

\[
\begin{bmatrix}
Y_{re} \\
Y_{im}
\end{bmatrix} =
\begin{bmatrix}
0 & -\sin \alpha \\
\sin \alpha & 0
\end{bmatrix}
\begin{bmatrix}
x_{re} \\
x_{im}
\end{bmatrix}. \tag{6}
\]

These considerations are in the basis of the WFTA and permit to transform (2) into (12). Therefore, WFTA requires only a semi-complex multiplier for multiplication with complex data and provides two times smaller hardware cost.

C. Fixed-Point Arithmetic

In exact arithmetic computations such a semi-complex multiplication would be easy to perform. However, in modern digital systems, the memory registers with finite capacity are used. Therefore, finite-precision arithmetic is used to represent values and perform computations.

Consider 2’s complement Fixed-Point Arithmetic [18], [19]. In this arithmetic, real numbers are represented as integers scaled by an implicit quantization factor. Such arithmetic benefits from fast bit operations. However, some accuracy concerns are related to such passage to finite precision.

Let \( r \) and \( i \) be the real and imaginary part of the complex constant \( c \). In our case we suppose that the input sequences are quantized such that \( r < 1 \) and \( i < 1 \), i.e. we have only fractional part to be represented. Therefore, if implemented with \( b \) bits, the real and imaginary parts are stored in machine as integers \( R, I \in \mathbb{Z} \cap [-2^{b-1}, 2^{b-1}-1] \) which are obtained by

\[
R = \lfloor 2^{b-1} \cdot \cos \alpha \rfloor, \quad I = \lfloor 2^{b-1} \cdot \sin \alpha \rfloor, \tag{7}
\]

where \( \lfloor \cdot \rfloor \) represents the rounding operation (round to nearest). Therefore, the Fixed-Point counterparts of \( r \) and \( i \) are \( R \cdot 2^{-b+1} \) and \( I \cdot 2^{-b+1} \) respectively. The quantization error \( \epsilon \) depends on the chosen rounding mode. In case of truncation, we obtain that

\[
-2^{-b+1} \leq \epsilon < 0. \tag{8}
\]

Another issue that inevitably follows the coefficient quantization is that the additions and multiplications are performed with finite number of bits as well, which may lead to computational errors. After being propagated, these errors may be significantly amplified. This issue will be addressed in Section IV.
III. PROPOSED ARCHITECTURE

The architecture of radix is obtained by direct implementation of the SFG from Fig. 1, which requires a number of hardware components equal to operations. Table I present the number of operations of each radix respectively. Here, under a trivial multiplication we mean multiplication by \( -j \) that can be simply performed by swapping the real and imaginary part of data. As result, it does not influence the overall hardware cost. It can be seen from the Table I that there is no single radix on which we could to map the rest of the radices: radix-3 is the only one to have a multiplier but it lacks two adders for the radix-4 architecture.

We, on the other side, propose to replace the classic WFTA implementation by the multiplierless version, which introduces additional adders instead of the multiplier, and enables us to map the radix-2/4/FFTs on it.

A. Multiplierless radix-3 WFTA architecture

A single multiplication by a constant is used in radix-3 WFTA architecture. It is possible to use shift-and-add circuit to efficiently implement the multiplication instead of general multiplier.

Generally, the canonical signed digit representation is used to reduce the number of non-zero digits with respect to the simple binary representation and, therefore the number of adders [20]. Further simplification is achieved by single constant multiplication techniques [21], [22]. They exploit the redundancy in the multiplication by single constant. Additionally, improvement in accuracy and complexity can be achieved by addition-aware coefficient quantization method [23].

Using these techniques proposed in [20]–[23], we obtain a number of candidate coefficients whose fractional bits range are tabulated in Table II.

The selection is based on specifications like adder cost and coefficient quantization error \( \epsilon \). It can be observed that three different techniques are required for different coefficient wordlengths in Table II:

- using 2 adders
  \[ C \cdot A = -1j \cdot ((A \cdot 2^3 - A) \cdot 2^4 - A) / 2^7, \]
  \( (9) \)

- using 3 adders:
  \[ C \cdot A = -1j \cdot ((A \cdot 2^3 - x) \cdot 2^7 - (A \cdot 2^3 + x)) / 2^{10}, \]
  \( (10) \)

- using 4 adders
  \[ C \cdot A = -1j \cdot ((A \cdot 2^3 - A) \cdot 2^2 - (A \cdot 2^3 - A) \cdot 2^3 \]
  \[ - (A \cdot 2^3 - A) \cdot 2^4 \]
  \[ - (A \cdot 2^3 - A) \cdot 2^{21}) / 2^{14}. \]
  \( (11) \)

TABLE I
NUMBER OF OPERATIONS OF SMALL POINT FFTS.

<table>
<thead>
<tr>
<th>Size</th>
<th>adds/subs</th>
<th>trivial mults</th>
<th>non-trivial mults</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radix-2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Radix-3</td>
<td>6</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Radix-4</td>
<td>8</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

TABLE II
DIFFERENT QUANTIZATIONS OF THE NON-TRIVIAL COEFFICIENT \( \frac{\sqrt{x}}{2} \) AND NUMBER OF ADDERS TO REALIZE THE MULTIPLIERLESS STRUCTURE.

<table>
<thead>
<tr>
<th>Fractional Bits</th>
<th>Coefficient</th>
<th>Adders</th>
<th>( \epsilon )</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>111 \times 2^{-7}</td>
<td>2</td>
<td>1.12 \times 10^{-3}</td>
</tr>
<tr>
<td>8</td>
<td>222 \times 2^{-8}</td>
<td>2</td>
<td>1.12 \times 10^{-3}</td>
</tr>
<tr>
<td>9</td>
<td>443 \times 2^{-9}</td>
<td>3</td>
<td>7.91 \times 10^{-4}</td>
</tr>
<tr>
<td>10</td>
<td>887 \times 2^{-10}</td>
<td>3</td>
<td>1.85 \times 10^{-4}</td>
</tr>
<tr>
<td>11</td>
<td>1774 \times 2^{-11}</td>
<td>3</td>
<td>1.85 \times 10^{-4}</td>
</tr>
<tr>
<td>12</td>
<td>3547 \times 2^{-12}</td>
<td>4</td>
<td>5.86 \times 10^{-5}</td>
</tr>
<tr>
<td>13</td>
<td>7094 \times 2^{-13}</td>
<td>4</td>
<td>5.86 \times 10^{-5}</td>
</tr>
<tr>
<td>14</td>
<td>14189 \times 2^{-14}</td>
<td>4</td>
<td>2.42 \times 10^{-6}</td>
</tr>
<tr>
<td>15</td>
<td>28378 \times 2^{-15}</td>
<td>4</td>
<td>2.42 \times 10^{-6}</td>
</tr>
</tbody>
</table>

Fig. 2. Multiplierless architecture for radix-3 (Multiplication \( [-\frac{\sqrt{x}}{2}]j \), 3 adders).

We selected the 10 \(-\) bit coefficient with 3 adders, which has the coefficient quantization error bounded by 1.85 \times 10^{-4}. Usually, the shifts are free of cost as they are hardwired. The final multiplierless architecture is illustrated on the Fig. 2.

B. Multiplierless unified radix-2/3/4 FFT

The multiplierless unified architecture is based on mapping the 2/3/4-radix into a single processing core. First, we design a multiplierless radix-3 FFT architecture. It contains enough computational resources (adders and substractors) to map the radix-2/4, the main challenge is to reduce the number of multiplexers. To obtain this, it is important to find common parts in the signal flow graphs that can be mapped without multiplexers. Further, multiplexer can be avoided by setting the unused inputs of the circuit to zero, which removes the unnecessary connections of the circuit. Using these techniques, a solution with only seven two-to-one multiplexers controlled with two control signals, has been designed. The resulting architecture is shown in Fig. 3. The input and output relations are tabulated in Table III, where the dashes denote “don’t care” conditions and 0 denotes that the input should be zeroed for proper operations. Finally, signals controlling the multiplexers are shown in Table IV, where dash denotes “don’t care” condition.

IV. ERROR ANALYSIS

Any practical finite precision implementation of the above discussed classical and new FFTs will suffer from finite-precision effects. Usually, for the finite-precision error analysis in signal processing domain a stochastic approach is
used: the quantization and computational errors are seen as a white uniformly distributed noise added to the signal. Such a statistical modeling, described in details by Widrow [24] or Constantinides [25], does not give precise values that the errors can take but provides the average and variance that describe the statistical dispersion of the error.

We, on the other hand, propose to perform accurate error-analysis while exploiting the properties of Fixed-Point Arithmetic. We will model the errors as intervals (tighter than we usually deduce with statistical approach) and investigate their propagation. In order to analyze the accuracy of any implementation, we need to fully specify the way the arithmetic operations are implemented in hardware. In our implementations we ensure the following:

- Multiplication: our Fixed-Point multiplier, denoted ⊗, guarantees that the product of two real numbers \( x \) and \( y \), that are represented with \( b_x \) and \( b_y \) bits respectively, is computed such that the output represented with \( b = \min\{b_x, b_y\} \) bits has its error is bounded by
  \[
  |x \cdot y - \text{exp}_\text{mul}(x, y)| < 2^{-b+1}
  \]

- Addition: we suppose that the input data is scaled such that no over- or under-flow occur in the computations; and since the operands will always be in the same format, no computational errors are introduced for the Fixed-Point addition.

### A. Classical WFTA radix-3 FFT

The classic WFTA radix-3 point FFT with a multiplier as shown in Fig. 2(c) can be represented as
\[
[X(0) ... X(N-1)]^T = B_3 \cdot B_3 \cdot W \cdot B_2 \cdot B_2 \cdot B_1 \cdot [x(0) ... x(N-1)]^T,
\]
where matrices \( B_i(3 \times 3) \) contain only trivial terms \((0, 1, -1)\) and \( W(3 \times 3) \) contains the non-trivial coefficient. Such computation can also be represented as an implicit system of linear equations:
\[
\begin{align*}
Jt & = Nx \\
y & = Lt
\end{align*}
\]
where
\[
J = \begin{pmatrix}
-1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & W & -1 & 0 \\
B_3 & -1 & 0 & 0
\end{pmatrix},
\]
\[
N = \begin{pmatrix}
B_1 \\
0 \\
0 \\
0
\end{pmatrix},
\]
\[
L = \begin{pmatrix}
0 & 0 & 0 & B_4
\end{pmatrix}
\]
and \( J \in \mathbb{C}^{12 \times 12}, N \in \mathbb{R}^{12 \times 3}, L \in \mathbb{R}^{3 \times 12}; \) and \( 1, 0 \in \mathbb{R}^{3 \times 3} \) are identity and zero matrices respectively.

**Quantization errors:** using the bound (8), we know that instead of \( J \) we actually implement the quantized to \( b \) bits matrix \( J_q := J - \Delta J \), where
\[
|\Delta J_{i,j}| \leq \begin{cases}
2^{-b+1}, & \text{if } J_{i,j} \text{ is the non-trivial coefficient} \\
0, & \text{otherwise}
\end{cases}
\]
Therefore, instead of system (13) we compute
\[
\begin{align*}
J_q \hat{t} & = N \hat{x} \\
\hat{y} & = L \hat{t}
\end{align*}
\]
**Multiplication errors:** consider the above assumptions concerning Fixed-Point multiplication and that the wordlength of non-trivial factor \( \sqrt{2} \) is at least equal to the the wordlength of the input. Then, instead of the exact vector \( t \), we actually compute \( t := t - \Delta t \), where
\[
|\Delta t| \leq \begin{cases}
2^{-b+1}, & \text{if } t_i \text{ contains non-trivial factor} \\
0, & \text{otherwise}
\end{cases}
\]
Therefore, the actually implemented output \( \hat{y} \) is computed with
\[
\begin{align*}
J_q \hat{t} & = N \hat{x} + \Delta t \\
\hat{y} & = L \hat{t}
\end{align*}
\]
Then, for a given input vector \( x \in \mathbb{R}^3 \), the implementation error of the radix-3 WFTA is bounded by:
\[
||y - \hat{y}|| \leq ||L (J_q^{-1} - J^{-1}) \hat{x}|x - L J_q \Delta t||,
\]
where \( ||\cdot||_\infty \) is some norm. Thanks to the sparse nature of \( \Delta J \) and \( \Delta t \), the error bound (19) can be simplified:

\[
||y - \hat{y}|| \leq ||Q \cdot x + p||,
\]
with
\[
Q = \begin{pmatrix}
0 & 0 & 2^{-b+1} \\
0 & -2^{-b+1} & 0 \\
0 & -2^{-b+1} & -2^{-b+1}
\end{pmatrix}, \quad p = \begin{pmatrix}
0 \\
-2^{-b+1} \\
2^{-b+1}
\end{pmatrix}
\] (21)

Matrix \( Q \) represents the impact of the coefficient quantization errors and vector \( p \) holds the impact of multiplication errors to the final output.

**Remark:** Consider the case, when the coefficient \( \frac{\sqrt{2}}{2} \) is quantized to \( b_1 \) bits, and \( b_2 \) is less than the wordlength \( b \) of the input/output of the FFT algorithm. Then, the bound \( p \) will change to
\[
\begin{pmatrix}
0 \\
-2^{-b_2} + 1 \\
2^{-b_2} + 1
\end{pmatrix}
\]

**B. Multiplierless radix-3 point FFT**

Similarly to the case of the WFTA radix-3 FFT, our multiplierless design can be represented as a system of linear equations (13) but with slightly different matrix \( J \). The difference between two designs is the use of shifts and adds technique instead of the multiplication. In order to show that the classic and proposed designs have equivalent behavior in terms of errors, consider following reasoning.

**Quantization errors:** the multiplierless architectures that we proposed in Section III-A are based on coefficients that have already been quantized (error bounds are given in Table II). Therefore, we obtain that modeling (15) holds.

**Multiplication errors:** if the additions in multiplierless architecture are performed with enough guard bits (it can be shown, that only 2 bits are required), then the error bound (17) is satisfied as well.

Therefore, our modeling of the error propagation stays the same and the new multiplierless design has an error bound equivalent to the implementation with a multiplier (or better if additions in the multiplierless multiplication are done with care).

**V. Conclusion**

In this paper, we have presented a multiplierless architecture for the radix-3 WFTA and provided the shift-andadders specifications for the non-trivial constant quantized to the 6 – 14 bits, which covers the majority of real-life applications. This radix-3 multiplierless architecture permitted us to map the radix-2 and 4 classical FFTs upon it. The proposed universal architecture requires only a few additional multiplexers controlled by two control signals. Finally, we proposed a new modeling for the Fixed-Point error analysis of the classical WFTA and showed that the new multiplierless design respects the same error bounds.

**Acknowledgment**

This work is partially supported by Tekes - the Finnish Funding Agency for Innovation under funding decision 40142/14 (StreamPro) in the FiDiPro program and by the ANR (French National Research Agency) grant ANR-13-INSE-0007-02 MetaLibm.

**References**


