A General Comparison Of Fft Algorithms

Manish Soni, Padma Kunthe

Abstract

A large number of FFT algorithms have been developed over the years, notably the Radix-2, Radix-4, Split-Radix, Fast Hartley Transform (FHT), Quick Fourier Transform (QFT), and the Decimation-in-Time-Frequency (DITF) algorithms. How these algorithms fare in comparison with each other is of considerable interest to developers of signal processing technology. In this paper, we present a general analysis and comparison of the aforementioned algorithms. The analysis of each algorithm includes the number of mathematical operations, computation time and memory requirements. The results indicate that the FHT is the overall best algorithm on all platforms, offering the fastest execution time and requiring reasonably small amounts of memory.

1. Introduction: The first major breakthrough in implementation of Fast Fourier Transform (FFT) algorithms was the Cooley-Tukey [1] algorithm developed in the mid-1960s, which reduced the complexity of a Discrete Fourier Transform from O(N^2) to O(N·logN). At that time, this was a substantial saving for even the simplest of applications. Since then, a large number of FFT algorithms have been developed. The Cooley-Tukey algorithm became known as the Radix-2 algorithm and was shortly followed by the Radix-3, Radix-4, and Mixed Radix algorithms [8]. Further research led to the Fast Hartley Transform (FHT), [2,3,4] and the Split Radix (SRFFT), [5] algorithms. Recently, two new algorithms have emerged: the Quick Fourier Transform (QFT), [6] and the Decimation-In-Time-Frequency (DITF), algorithm [7]. In this paper we provide a comparison of several contemporary FFT algorithms. The criteria used are the operations count, memory usage and computation time. We chose the following algorithms for our analysis: Radix-2 (RAD2), Radix-4 (RAD4), SRFFT, FHT, QFT and DITF.

2. Review of FFT algorithms: The basic principle behind most Radix-based FFT algorithms is to exploit the symmetry properties of a complex exponential that is the cornerstone of the Discrete Fourier Transform (DFT). These algorithms divide the problem into similar sub-problems (butterfly computations), and achieve a reduction in computational complexity. All Radix algorithms are similar in structure differing only in the core computation of the butterflies. The FHT differs from the other algorithms in that it uses a real kernel, as opposed to the complex exponential kernel used by the Radix algorithms. The QFT postpones the complex arithmetic to the last stage in the computation cycle by separately computing the Discrete Cosine Transform (DCT), and the Discrete Sine Transform (DST). The DITF algorithm uses both the Decimation-In-Time (DIT), and Decimation-In-Frequency (DIF), frameworks for separate parts of the computation to achieve a reduction in the computational complexity.

Radix-2 Decimation in Frequency Algorithm: The RAD2 DIF algorithm is obtained by using the divide-and-conquer approach to the DFT problem. The DFT computation is initially split into two summations, one of which involves the sum over the first data points and the other over the next data points, resulting in[11,12].

*Research Scholor M.Tech (Digital Comm.), TIT,College,Bhopal
**Asst.Prof.,Head EC Deptt. TIT,College,Bhopal
The same computational procedure can be repeated through decimation of the N/2-point DFTs X(2k), and X(2k+1). The entire process involves \( v = \log_2 N \) stages with each stage involving N/2 butterflies. Thus the RAD2 algorithm involves \( N/2 \cdot \log_2 N \) complex multiplications and \( N \cdot \log_2 N \) complex additions, or a total of \( 5N \cdot \log_2 N \) floating point operations. Observe that the output of the whole process is out-of-order and requires a bit reversal operation to place the frequency samples in the correct order.

2.2 Radix-4 Algorithm: The RAD4 algorithm is very similar to the RAD2 algorithm in concept. Instead of dividing the DFT computation into halves as in RAD2, a four-way split is used. The N-point input sequence is split into four subsequences, \( x(4n) \), \( x(4n+1) \), \( x(4n+2) \), and \( x(4n+3) \), where \( n=0,1,\ldots,N/4-1 \). Then,

\[
X(k) = \sum_{n=0}^{N-1} x(n) \cdot W_N^{kn} + \sum_{n=N/4}^{N-1} x(n) \cdot W_N^{kn} + \sum_{n=3N/4}^{2N-1} x(n) \cdot W_N^{kn} + \sum_{n=N/2}^{N-1} x(n) \cdot W_N^{kn} \tag{5}
\]

Setting

\[
F(l, q) = \sum_{m=0}^{N-1} x(l, m) \cdot W_{N/4}^{mq} \tag{6}
\]

\[
X(p, q) = X \left[ \frac{N}{4} \cdot p + q \right], \tag{7}
\]

and,

\[
X(lm) = x(4m+1), \text{ where } l, p = 0, 1, 2, 3 \text{ & } m, q = 0, 1, \ldots, N/4-1. \tag{8}
\]

the matrix formulation of the butterfly becomes
The decimation process is similar to the RAD2 algorithm, and uses \( v = \log_4 N \) stages, where each stage has \( N/4 \) butterflies. The RAD4 butterfly involves 8 complex additions and 3 complex multiplications, or a total of 34 floating point operations. Thus, the total number of floating point operations involved in the RAD4 computation of an \( N \)-point DFT is \( 4.25 \log_2 N \), which is 15% less than the corresponding value for the RAD2 algorithm.

2.3. Split-Radix Algorithm: Standard RAD2 algorithms are based on the synthesis of two half-length DFTs and similarly RAD4 algorithms are based on the fast synthesis of four quarter-length DFTs. The SRFFT algorithm is based on the synthesis of one half-length DFT together with two quarter-length DFTs. This is possible because, in the RAD2 computations, the even-indexed points can be computed independent of the odd indexed points. The SRFFT algorithm uses the RAD4 algorithm to compute the odd numbered points. Hence, the \( N \)-point DFT is decomposed into one \( N/2 \)-point DFT and two \( N/4 \)-point DFTs.

An \( N \)-point DFT is obtained by successive use of these decompositions. Here we treat the computational process as a RAD2 algorithm with the unnecessary intermediate DFT computations eliminated. An analysis of the butterfly structures [15] for the SRFFT algorithm reveals that approximately \( 4N \log_2 N \) computations are required as compared to \( 4.25 \log_2 N \) for RAD4 and \( 5N \log_2 N \) for RAD2 algorithms.

2.4. Fast Hartley Transform: The main difference between the DFT computations previously discussed and the Discrete Hartley Transform (DHT), is the core kernel [2,14]. For the DHT, the kernel is real unlike the complex exponential kernel of the DFT. The DHT coefficient is expressed in terms of the input data points as [13]

\[
X(k) = \sum_{n=0}^{N-1} x(n) \cdot \left[ \cos \left( \frac{2\pi nk}{N} \right) + \sin \left( \frac{2\pi nk}{N} \right) \right] \quad (14)
\]

This results in the replacement of complex multiplications in a DFT by real multiplications in a DHT. For complex data, each complex multiplication in the summation requires four real
multiplications and two real additions using the DFT. For the DHT, this computation involves only two real multiplications and one real addition. There exists an inexpensive mapping of coefficients from the Hartley domain to the Fourier domain, which is required to convert the output of a DHT to the traditional DFT coefficients. Equation (15), relates the DFT coefficients to the DHT coefficients for an N-point DFT computation.

\[
\begin{align*}
\text{Re}(DFT(k)) &= \frac{DHT(k) + DHT(N-k)}{2} \\
\text{Im}(DFT(k)) &= \frac{DHT(k) - DHT(N-k)}{2}
\end{align*}
\]

(15)

The FHT evolved from principles similar to those used in the RAD2 algorithm to compute DHT coefficients efficiently[13]. It is intuitively simpler and faster than the FFT algorithms as the number of computations reduces drastically when we replace all complex computations by real computations. Similar to other recursive Radix algorithms, the next higher order FHT can be obtained by combining two identical preceding lower order FHTs. In fact all Radix-based algorithms used in FFT implementations can be applied to FHT computations [16]. For N=2, the Hartley transform can be represented in a matrix form as

\[
\begin{bmatrix}
\mathcal{X}(0) \\
\mathcal{X}(1)
\end{bmatrix} =
\begin{bmatrix}
1 & 1 \\
1 & -1
\end{bmatrix}
\begin{bmatrix}
\mathcal{x}(0) \\
\mathcal{x}(1)
\end{bmatrix}
\]

(16)

Following a similar procedure for, we get the matrix formulation,

\[
\begin{bmatrix}
\mathcal{X}(0) \\
\mathcal{X}(1) \\
\mathcal{X}(2) \\
\mathcal{X}(3)
\end{bmatrix} =
\begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & -1 & -1 & -1 \\
1 & -1 & 1 & -1 \\
1 & -1 & -1 & 1
\end{bmatrix}
\begin{bmatrix}
\mathcal{x}(0) \\
\mathcal{x}(1) \\
\mathcal{x}(2) \\
\mathcal{x}(3)
\end{bmatrix}
\]

(17)

which can be easily transformed to

\[
\begin{bmatrix}
\mathcal{X}(0) \\
\mathcal{X}(1) \\
\mathcal{X}(2) \\
\mathcal{X}(3)
\end{bmatrix} =
\begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & -1 & -1 & -1 \\
1 & -1 & 1 & -1 \\
1 & -1 & -1 & 1
\end{bmatrix}
\begin{bmatrix}
\mathcal{x}(0) \\
\mathcal{x}(1) \\
\mathcal{x}(2) \\
\mathcal{x}(3)
\end{bmatrix}
\]

(18)

A closer look at this matrix product and a comparison with the matrix for N=2 reveals that the matrix for N=4 is composed of sub-matrices of the form of the matrix for N=2. Thus a DHT of order 4 can be computed directly from a DHT of order 2. This idea can be extended to any order which is a power of 2 [4]. It is also worth noting that the Hartley Transform is a bilateral transform, i.e. the same functional form can be used for both the forward and inverse transforms. This is an added advantage of the FHT over other FFT algorithms.

2.5. Quick Fourier Transform: We have seen that the Radix-based algorithms exploit the periodic properties of the cosine and sine functions. In the Quick Fourier Transform (QFT), algorithm, the symmetry properties of these functions are used to derive an efficient algorithm.
We define an $N+1$-point DCT as

$$X_{\text{DCT}}(k) = \sum_{n=0}^{N} x(n) \cos \left( \frac{\pi nk}{N} \right), \quad k = 0, 1, \ldots, N$$  \hspace{1cm} (20)$$

An $N-1$-point DST can also be similarly defined as

$$X_{\text{DST}}(k) = \sum_{n=1}^{N-1} x(n) \sin \left( \frac{\pi nk}{N} \right), \quad k = 0, 1, \ldots, N$$  \hspace{1cm} (21)$$

We can divide an $N$-point input sequence into its even and odd parts as

$$x_e(0) = x(0) \quad  \quad  x_o(k) = x(k) + x(N-k), \quad k = 1, 2, \ldots, N/2 - 1 \text{ and}$$

$$x_e(N/2) = x(N/2)$$  \hspace{1cm} (22)$$

Using the above sequences and properties in Equation 19 we can define an $N$-point DFT as

$$X(k) = X_{\text{DCT}}(k) - jX_{\text{DST}}(k), \quad k = 1, 2, \ldots, N/2 - 1$$  \hspace{1cm} (24)$$

In order to derive a recursive formulation of DCT and DST computations, we define a new sequence $x_e$ as

$$x_e(k) = x(k) + x(N - k), \quad k = 1, 2, \ldots, N/2 - 1$$  \hspace{1cm} (25)$$

Also, the $N/2$th point of this sequence is the same as that of the original sequence. Thus we can formulate the recursive DCT for the even numbered points as

$$DCT(2k, N+1, x_e) = DCT\left(k, \frac{N}{2}+1, x_e\right)$$

where $k = 1, 2, \ldots, N/2 - 1$  \hspace{1cm} (26)$$

$$X(0) = X_{\text{DCT}}(0) \text{ and } X(N/2) = X_{\text{DCT}}(N/2)$$  \hspace{1cm} (27)$$

We can define a recursive equation for the odd DCT points using a new sequence $x_o$ defined as

$$x_o(k) = \frac{x(k) - x(N-k)}{2\cos \left( \frac{\pi k}{N} \right)}, \quad k = 1, 2, \ldots, N/2 - 1$$  \hspace{1cm} (28)$$

Then,

$$DCT(2k + 1, N + 1, x_e) = DCT\left(k, \frac{N}{2} + 1, x_o\right) + DCT\left(k + 1, \frac{N}{2} + 1, x_o\right), \quad k = 0, 1, \ldots, N$$  \hspace{1cm} (29)$$
A similar recursive formulation can be derived for the DST using symmetry properties of the sine function which results in
\[ DST(2k, N-1, \chi) = DST(k, \frac{N-1}{2}, \chi) \] and
\[ DST(2k+1, N-1, \chi) = DST(k, \frac{N-1}{2}, \chi) + DST(k+1, \frac{N-1}{2}, \chi) \]
where \( k = 1, 2, \ldots \frac{N}{2}-1 \).
Since the complex operations occur only in the last stage of the computation where the DCT and DST are combined using Equation 24, the QFT is well suited for operation on real data. The number of operations required to perform an \( N \)-point QFT is
\[ \frac{11}{2}N \cdot \log N - \frac{27}{4}N + 2 \] [7]. This, however, does not include the cost of computing the odd and even parts of the data sequence at each stage of the computation.

2.6. Decimation-In-Time-Frequency (DITF), Algorithm: The DITF algorithm is based on the observation that in a DIF implementation of a RAD2 algorithm, most of the computations (especially complex multiplications), are performed during the initial stages of the algorithm. In the DIT implementation of the RAD2 algorithm, the computations are concentrated towards the final stages of the algorithm. Thus, starting with the DIT implementation and then shifting to the DIF implementation at some transition stage intuitively seems to be a computation saving process. Equations (3), and (4), define the DIF RAD2 computation. The DIT RAD2 computation is defined as
\[ X(k) = \sum_{n=0}^{\frac{N}{2}-1} x(2n).W_N^{k} + \sum_{n=0}^{\frac{N}{2}-1} x(2n+1).W_N^{k} \]
\[ (31) \]
Note that the first summation in the above equation is the \( N/2 \)-point DFT of the sequence comprised of the even-numbered points of the original sequence and the second summation is the \( N/2 \)-point DFT of the sequence comprised of the odd numbered points of the original sequence. The transition stage consists of a conversion from the DIT coefficients to the DIF coefficients,
\[ DIF(k) = W_N^{PQ} \cdot DIT(k) \] [32]
where \( p \) is the index of the set to which \( k \) belongs and \( q \) is the position of \( k \) in that set. The indices of each set need to be bit reversed. The total number of real multiplications involved in the DITF computation is \( 2N \cdot \log N - 10N + 8N/2s + 8.2s - 8 \), where \( s \) is the transition stage. On minimizing this expression, we get the optimal transition stage for minimum number of multiplications as
\[ \frac{\log N}{2} \].

3. Benchmarking criteria: Most preceding FFT complexity studies have been conducted on special purpose hardware such as digital signal processing (DSP), chips [9,10]. Typically, the primary benchmarking criteria have been the number of mathematical operations (multiplications and additions), and/or the overall computation speed. The efficiency of an algorithm is most influenced by the arithmetic complexity, usually expressed in terms of a count of real multiplications and additions. However, on general purpose computers this is not a very good
benchmark and other factors need to be considered as well. For instance, the issue of memory usage is very important for memory constrained applications.

3.1. Number of Computations: Since many general purposes CPUs have significantly different speeds on floating point and integer operations, we decided to individually account for floating point and integer arithmetic. It is a well known fact that most new architectures compute floating point operations more efficiently than integer operations [19,21]. Also, most indexing and loop control is done using integer arithmetic. Therefore the integer operations count directly measures the cost of indexing and loop control. Many FFT algorithms require a large number of division-by-two operations which is efficiently accomplished by using a binary shift operator. To account for this common operation, we include a count of binary shifts in our benchmarks.

3.2. Computation Speed: In most present-day applications for general purpose computers, with easy availability of faster CPUs and memory not being a primary constraint, the fastest algorithm is by far treated as the best algorithm. Thus, a common choice to rank order algorithms is by their computation speed.

3.3. Memory Usage: One of the classic trade-offs seen in algorithm development is that of memory usage versus speed. In most portable signal processing applications, the FFT is a core computational component. However, few applications can afford a large memory space for evaluating FFTs. While memory usage is important for specification of hardware, memory accesses also account for a significant portion of computation time. This is attributed to cache misses, swapping and other paging effects. These effects are more prominent when computing higher order FFTs (typically over 4K points). These observations prompted us to include memory usage as one of the yardsticks in judging the effectiveness of the various FFT algorithms.

4. Benchmarking results and analysis: Each of the algorithms was implemented under a common framework using common functions for operations such as bit-reversal and lookup table generation so that differences in performance could be attributed solely to the efficiency of the algorithms. Following this, we comprehensively benchmarked each algorithm according to the criteria discussed in the previous section.

4.1. Computation Speed: Computation speed is typically the most prominent aspect of an FFT algorithm in current DSP applications. The computation speed of an algorithm for large data sizes can often be heavily dependent on the clock speed, RAM size, cache size and the operating system. Hence, these factors must be taken into account. We evaluated that the computation time of the worst algorithm (DITF), is more than three times greater than that of the best algorithm (FHT). It has been consistently observed in our benchmarks that the FHT is the most efficient algorithm in terms of computation speed. Table 1 shows the variation in performance of these algorithms as a function of the FFT order. The performance is clearly affected by the amount of RAM and cache. As expected, the effect is more pronounced for higher order FFTs where cache misses become common. The performance of the CPUs on floating point operations versus integer operations is significant as well.

4.2. Number of Computations: The number of arithmetic computations has been the traditional measure of algorithmic efficiency. The numbers of operations required by each algorithm for a 1024-point real DFT are displayed in Table 2. We observe that the faster algorithms require performing a smaller number of computations. However, there is a trade-off between integer operations and floating point operations. Savings in floating point operations can be achieved at the cost of increasing the number of integer operations. An example of this is seen in the excessive number of integer additions in the QFT. In the QFT implementation, the DCT and DST
recursions are implemented by accessing pointers in a common workspace. This results in the large number of integer operations. The large numbers of operations for the DITF algorithm are attributed to the bit-reversal process at various stages of the computation. This aspect seems to have been overlooked in previous evaluations [7,8]. Overall, the FHT and the SRFFT are the best in terms of effectively using computations, which translates to greater computation speed. The main drawback of the FHT is that the complex FHT is computed via two real FHT computations. The QFT also uses a similar methodology. The number of computations doubles when moving from real data to complex data using these algorithms. The corresponding change for the other algorithms is insignificant.

4.3. Memory Usage: One of the key issues in portable applications is memory usage. Table 3 shows the memory usage profile of different algorithms for a 1024 point FFT. We see from Table 3 that the RAD2 algorithm is the most memory efficient algorithm, and the QFT is the least. In the case of the QFT, this is due to the large work space required to perform the recursions in the DCT and the DST algorithms. The FHT is the most inefficient in terms of the executable size. Notice that, as was expected, the executable size is a good measure of the complexity of the algorithm with the FHT being the most complex and the RAD2 the least complex algorithm.

5. Conclusions: The existence of an abundance of algorithms for FFT computations and an even greater number of their implementations calls for a comprehensive benchmark which teases out the implementation-specific differences and compares the algorithms directly. We have tried to achieve this objective by implementing algorithms in a very consistent framework. Our results indicate that the overall best algorithm for DFT computations is the FHT algorithm. This has been, and will likely continue to be, a point of argument for many years [17, 18, 20, 22, 23]. Another feature in favor of the FHT is its bilateral formulation. Unlike DFT algorithms, FHT has the same functional form for both its forward and inverse transforms. The FHT is the fastest algorithm on all platforms with a reasonable dynamic memory requirement. However, it is the most inefficient in terms of static memory usage (measured in terms of the executable size). If an FFT algorithm needs to be chosen solely on the basis of static memory requirements, the RAD-2 algorithm is the still the best, owing to its simple implementation. The SRFFT and the FHT are comparable in terms of the number of computations and are the most efficient.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>FFT order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>16</td>
</tr>
<tr>
<td>RAD2</td>
<td>20</td>
</tr>
<tr>
<td>RAD4</td>
<td>20</td>
</tr>
<tr>
<td>SRFFT</td>
<td>20</td>
</tr>
<tr>
<td>FHT</td>
<td>20</td>
</tr>
<tr>
<td>QFT</td>
<td>20</td>
</tr>
<tr>
<td>DITF</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 1: Computation time (in microseconds) of various algorithms.
<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Float Adds</th>
<th>Float Mults</th>
<th>Integer Adds</th>
<th>Integer Mults</th>
<th>Binary Shifts</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAD2</td>
<td>14336</td>
<td>20480</td>
<td>19450</td>
<td>2084</td>
<td>1023</td>
</tr>
<tr>
<td>RAD4</td>
<td>8960</td>
<td>14336</td>
<td>12902</td>
<td>3071</td>
<td>277</td>
</tr>
<tr>
<td>SRFFT</td>
<td>5961</td>
<td>5522</td>
<td>12664</td>
<td>2542</td>
<td>1988</td>
</tr>
<tr>
<td>FHT</td>
<td>7420</td>
<td>8841</td>
<td>3235</td>
<td>2048</td>
<td>12</td>
</tr>
<tr>
<td>QFT</td>
<td>9026</td>
<td>2560</td>
<td>29784</td>
<td>1048</td>
<td>144</td>
</tr>
<tr>
<td>DITF</td>
<td>14400</td>
<td>17664</td>
<td>20333</td>
<td>1076</td>
<td>1074</td>
</tr>
</tbody>
</table>

Table 2: Number of computations involved in computing a 1024-point FFT.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Memory Usage (Bytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAD2</td>
<td>72240</td>
</tr>
<tr>
<td>RAD4</td>
<td>72536</td>
</tr>
<tr>
<td>SRFFT</td>
<td>72508</td>
</tr>
<tr>
<td>FHT</td>
<td>72652</td>
</tr>
<tr>
<td>QFT</td>
<td>122072</td>
</tr>
<tr>
<td>DITF</td>
<td>78532</td>
</tr>
</tbody>
</table>

Table 3: Memory usage in computing a 1024-point FFT.
References:
