<table>
<thead>
<tr>
<th>Title</th>
<th>Error analysis and complexity optimization for the multiplier-less FFT-like transformation (ML-FFT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Tsui, KM; Chan, SC; Tse, KW</td>
</tr>
<tr>
<td>Issued Date</td>
<td>2004</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10722/46429">http://hdl.handle.net/10722/46429</a></td>
</tr>
<tr>
<td>Rights</td>
<td>Creative Commons: Attribution 3.0 Hong Kong License</td>
</tr>
</tbody>
</table>
ERROR ANALYSIS AND COMPLEXITY OPTIMIZATION FOR THE MULTIPLIER-LESS FFT-LIKE TRANSFORMATION (ML-FFT)
K. M. Tsui, S. C. Chan and K. W. Tse
Department of Electrical and Electronic Engineering,
The University of Hong Kong, Pokfulam Road, Hong Kong.

ABSTRACT
This paper studies the effect of the signal round-off errors on the accuracies of the multiplier-less Fast Fourier Transform-like transformation (ML-FFT). The idea of the ML-FFT is to parameterize the twiddle factors in the conventional FFT algorithm as certain rotation-like matrices and approximate the associated parameters inside these matrices by the sum-of-power-of-two (SOPOT) or canonical signed digits (CSD) representations. The error due to the SOPOT approximation is called the coefficient round-off error. Apart from this error, signal round-off error also occurs because of insufficient wordlengths. Using a recursive noise model of these errors, the minimum wordlength to realize the ML-FFT subject to the prescribed output bit accuracy can be obtained using a random search algorithm. A design example is given to demonstrate the effectiveness of the proposed approach.

I. INTRODUCTION
The Discrete Fourier Transform (DFT) is an important tool in digital signal processing [1]. A treasure of fast algorithms such as the Cooley-Tuky Fast Fourier Transform (FFT) and the prime factor algorithm (PFA) FFT are available to compute efficiently DFT of different lengths. Recently, the efficient realization of the multiplier-less FFT based on the integer [2,3] or SOPOT representation [4], and its extension to the multiplier-less sincosoidal transforms [5] have been proposed. The main objective is to avoid the expensive general-purpose multipliers which are replaced with limited number of shifters and adders. However, this approximation will unavoidably introduce errors which are referred to the coefficient round-off errors. Fortunately, as proposed in [4], tradeoffs between the arithmetic complexities and the output accuracies can be made so that the minimum arithmetic complexities can be obtained for different applications, which require different degree of the error tolerance. Due to finite wordlength of internal representation, another source of error, called signal round-off error [1], occurs when rounding is performed for the intermediate data after complex multiplication with the twiddle factor. Moreover, overflow can occur due to insufficient internal wordlength when fixed-point arithmetic is used. Unfortunately, most design methods for the multiplier-less FFT only focus on the effect of the coefficient round-off errors. In order to satisfy the prescribed output accuracy, one usually employs fixed but rather long wordlength for all intermediate data, which means increased hardware complexity. Therefore, it is necessary to design a general model to determine the minimum hardware complexity, subject to a given output accuracy.

In this paper, we propose a new recursive round-off noise model for computing the output bit accuracies of the ML-FFT under finite wordlength effect. Without loss of generality, the decimation-in-time (DIT) radix-p ML-FFT is used as an example. The noise sources due to the rounding operations performed after multiplications are first identified at each stage, based on the structure of the DIT radix-p FFT, where the size of the transformation N is the integer power of p. For each output point at any stage, its noise powers are determined statistically by its associated noise sources, using the commonly used uncorrelated white noise model. Together with the noise powers coming from the previous stage, the total noise powers of all the output points can be calculated, and propagate to the next stage. Eventually, the final output bit accuracy of each output point can be obtained by summing the total noise powers accumulated at this output point. Using these results, the internal wordlength of each intermediate data can then be optimized subject to prescribed output bit accuracy using a random search algorithm [8,9]. As an illustration, the number of adder cells and registers used, which is related to the exact wordlength used for each intermediate data, is chosen as a measure of the hardware complexity. Design result shows that our proposed approach can efficiently determine the minimum hardware complexity subject to prescribed output bit accuracy. The rest of this paper is organized as follows: Section II describes the ML-FFT algorithm based on the DIT radix-p FFT. Section III is devoted to the error analysis of the ML-FFT and the wordlength determination method. A design example demonstrating the effectiveness of the proposed approach is given in Section VI. Finally, conclusion is drawn in Section VII.

II. THE ML-FFT ALGORITHM
A. The decimation-in-time (DIT) radix-p FFT
The discrete Fourier transform (DFT) of an N-point sequence \( x(n) \) is given by:
\[
X(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi nk/N}, \quad k = 0,1,...,N-1
\]
where \( W_N = e^{-j2\pi/N} \) and \( i = \sqrt{-1} \). Consider the general formula of the DIT radix-p FFT as follow:
\[
X[k+r(N/p)] = \sum_{n=0}^{N-1} x(n)[W_p]^r, \quad r = 0,1,...,p-1
\]
for \( k = 0,1,...,N/p-1 \) and \( r = 0,1,...,p-1 \). Figure 1 shows the block diagram of the DIT radix-p FFT for computing a DFT of length \( N = p^r \). Using the above decomposition, the DFT can be reduced successively to \( N/p \)-point DFTs. In general, this process can be repeated \( m \) times and therefore there are totally \( m \) stages in the implementation of the DFT.

B. Multiplier-less realization
The ML-FFT algorithm [4] approximates the twiddle factor multiplications \( W_p^k \) and \( W_p^r \) in (2-2) by representing the coefficients of a certain factorization of the rotation using SOPOT coefficients. To start with, let \( c = (x+i\cdot y) \) be any complex number. The multiplication of \( c \) with \( \exp(-i\theta) \), \( p = c \cdot \exp(-i\theta) \) can be written as:
\[
\begin{bmatrix}
    p_r \\
    p_i
\end{bmatrix} = \begin{bmatrix}
    \cos\theta & \sin\theta \\
    -\sin\theta & \cos\theta
\end{bmatrix} \begin{bmatrix}
    x \\
    y
\end{bmatrix} = \begin{bmatrix}
    1 & 0 \\
    0 & -1
\end{bmatrix} R_c \begin{bmatrix}
    x \\
    y
\end{bmatrix}
\]
where \( R_c = \begin{bmatrix}
    \cos\theta & \sin\theta \\
    -\sin\theta & -\cos\theta
\end{bmatrix} \) is the rotation-like matrix and \( \theta = 2\pi k/N \) for \( W_p^k \) or \( \theta = 2\pi r/p \) for \( W_p^r \). However, \( \cos\theta \) and \( \sin\theta \) in (2-3) are not suitable for directly conversion to the SOPOT representation since the inverse of \( R_c \) cannot in general be expressed in terms of SOPOT coefficients. To cope with this, \( R_c \) is re-written as follows:
\[
R_c = \begin{bmatrix}
    1 & -\tan(\theta/2) \\
    0 & 1
\end{bmatrix} \begin{bmatrix}
    1 & 0 \\
    0 & 1
\end{bmatrix} \begin{bmatrix}
    1 & \tan(\theta/2) \\
    0 & 1
\end{bmatrix} = R_c'.
\]
Since the forward and inverse of the matrices involve the same set of coefficients, i.e. $\sin \theta$ and $\tan(\theta/2)$. They can be quantized to the SOPOT coefficients to from

$$R_{q} = \begin{bmatrix} 1 & -\beta_{q} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

where $\alpha_{q}$ and $\beta_{q}$ are respectively the SOPOT approximations to $\sin \theta$ and $\tan(\theta/2)$ having the form:

$$z_{q} = \sum_{r} a_{q} \beta_{q}^{r}, \quad (2-6)$$

where $a_{q} \in \{-1,1\}, b_{q} \in \{-r,\ldots,-1,0,\ldots,r\}; r$ is the range of the coefficients and $t$ is the number of terms used in each coefficient. Using these results, the number of SOPOT terms can then be optimized using the random search algorithm [7] subject to the specified errors between the candidate transform and its ideal counterpart. These errors due to the SOPOT approximation are called coefficient round-off errors, which can be reduced by using more SOPOT terms. Interested readers can refer to [4] for more details. In this section, we shall present the analysis of another noise source called signal round-off error, which will also affect the output accuracy of the ML-FFT.

### III. ROUND-OFF ANALYSIS OF THE ML-FFT

#### A. Signal round-off analysis

Signal round-off errors occur due to rounding of the intermediate signal after multiplications. Since the exact round-off errors are difficult to analyze, they are usually treated as uncorrelated white noises. For rounding operation, the quantization noise will have zero mean with variance $\sigma$ equal to $\Delta^2/12$, where $\Delta$ is the quantization step-size. In other words, the variance is determined by the number of fractional bits that is retained after multiplication. In fixed-point arithmetic, each intermediate signal can be represented in the form of $<I/F>$, where $I$ is the number of integer bits including the sign bit and $F$ is the number of fractional bits. In general, if $F$ bits are rounded to $B$ bits, where $B < F$, then the noise variance will be given by:

$$\sigma = \Delta^2/12 = 2^{-2(\text{frac} - 1)}/12. \quad (3-1)$$

Without loss of generality, consider the computation of an $N$-point $(N = p^m)$ FFT using the DIT radix-$p$ FFT algorithm as shown in the figure 2.

$$\epsilon_{\text{round-off}}(k) \text{ for } k = 0,1,\ldots, N/p - 1 \text{ and } j = 0,1,\ldots, p - 1 \text{ are the signal round-off noises introduced at the } j\text{-th output of the } j/(N/p)\text{-point DFT. The superscript } (k) \text{ indicates the } k\text{-th stage of the FFT structure, for } v = 0,1,\ldots,m. \text{ Hence, } v = 0 \text{ and } v = m \text{ correspond to the input and the output of the radix-$p$ FFT respectively. That is } X(a) = X^{(a)}(a) \text{ and } X^*(a) = X^{*(a)}(a) \text{ for } a = 0,1,\ldots,N-1. \text{ In general, } \epsilon_{\text{round-off}}^{(k)} \text{ are white and identically distributed, by symmetry consideration.}

$$n^{(k)}(k)'s \text{ are the additional signal round-off noises introduced by the multiplication with the "approximated" twiddle factors } W_{k}^{p}. \text{ They are modeled as zero mean white Gaussian noise with noise power depending on the finite word length after performing the multiplications. It should be noted that details for the other noise sources due to the multiplication with the twiddle factors } W_{k}^{p} \text{ are omitted in the figure 2 for simplicity. These errors only exist in the radix-8 or higher radices FFT algorithm. In the radix-2 and radix-4 FFT algorithms, the twiddle factors } W_{k}^{p} \text{ are } \pm 1 \text{ or } \pm i \text{ only, so the DFT does not require any multiplications and there is no rounding error in their implementation. In the rest of this section, } p \text{ is assumed to be equal to 2 or 4. However, it can easily be generalized to higher radices or split-radix FFT algorithms. Next we will discuss the determination of } n^{(k)}(k).$$

The round-off noise introduced by the rotation-like matrix $R_{q}$ in (2-4) can be computed as in the figure 3. If rounding is performed after each multiplication, three additive noise sources will be introduced as shown in the figure. Let $\sigma_{\epsilon}$ be the variances of the noise sources $r_{i}$ for $i = 0,1,2$. Assume $\sigma_{\epsilon}$ are uncorrelated, white and zero mean, it follows that the output noise variances are given by:

$$\sigma_{\epsilon}^2 = (1 + \sin^2 \theta \tan^2 (\theta/2))\sigma_{\epsilon}^2 + \left(\tan^2 (\theta/2)\right)\sigma_{\epsilon}^2 + k_{2}^2 \quad (3-2)$$

$$\sigma_{\epsilon}^2 = (\sin^2 \theta)\sigma_{\epsilon}^2 + \sigma_{\epsilon}^2. \quad (3-3)$$

Thus, the noise variance of $n^{(k)}(k)$ can be written as follows:

$$VAR[n^{(k)}(k)] = \sigma_{\epsilon}^2 + \sigma_{\epsilon}^2. \quad (3-4)$$

(3-2) and (3-3) can also apply to the ML-FFT with $\sin \theta$ replaced by their SOPOT approximations.

By interchanging the summation signs in (2-2), we can see that the real and imaginary parts of the output signal increase by no more than a factor of $p$ from stage to stage, assuming that both real and imaginary part of the output signal are less than one. Therefore, to avoid overflow of the immediate data, the outputs of the $p$-point DFT are usually scaled by a factor of $1/p$.

However, another noise sources $s^{(k)}(q)$, where $q = k + r(N/p)$ due to the scaling is introduced. More precisely, if all the precision in multiplying the SOPOT coefficients are retained and only the real part of the final result is rounded to $L_{p}$ bits, then the noise source for $Re(X^{(k)}(q))$ have variance equal to

$$VAR[Re(X^{(k)}(q))] = \frac{\Delta^2}{12} \frac{(3-5)}{12}. \quad (3-5a)$$

Similarly, when the imaginary part of the final part is rounded to $L_{p}$ bits, then the variance of the noise source for $Im(X^{(k)}(q))$ is given by:

$$VAR[Im(X^{(k)}(q))] = \frac{\Delta^2}{12} \frac{(3-5b)}{12}. \quad (3-5b)$$

Hence, we have

$$VAR[n^{(k)}(k)] = VAR[Re(X^{(k)}(q))] + VAR[Im(X^{(k)}(q))]. \quad (3-6)$$

Assuming that the round-off noises are uncorrelated. For other radices, the DFT might introduce additional noise sources, which depend on the exact implementation. Another point worth mentioning is that another scaling factor of $N$ at the final output is needed in order to obtain the correct DFTs, but this matter would not be taken into account for our noise model.

$$e^{(\text{round-off})}(k) \text{, } n^{(k)}(k) \text{ and } s^{(k)}(k) \text{ together constitute signal round-off noise at each output } X(q) \text{ and is denoted by } e^{(\text{round-off})}(q), \text{ which is also zero mean and Gaussian distributed. Assuming that the various noise sources are uncorrelated, the variance of } e^{(\text{round-off})}(q) \text{ at the } m\text{-th stage is given by:}

$$VAR[e^{(\text{round-off})}(q)] = \frac{1}{p^2} \sum_{j=0}^{p-1} VAR[e^{(\text{round-off})}(k)] \left| \hat{W}_{j}\hat{W}_{j}^{*} \right| + \frac{1}{p} \sum_{j=0}^{p-1} VAR[n^{(k)}(k)] \left| \hat{W}_{j}^{2} \right| + VAR[s^{(k)}(q)]. \quad (3-7)$$

Further, if we assume $e^{(\text{round-off})}(k)$ are identically distributed with variance $VAR[e^{(\text{round-off})}(k)]$, then (3-7) simplifies to

$$VAR[e^{(\text{round-off})}(q)] = \frac{VAR[e^{(\text{round-off})}(k)]}{p^2} \sum_{j=0}^{p-1} \left| \hat{W}_{j}\hat{W}_{j}^{*} \right| + \frac{VAR[n^{(k)}(k)]}{p} \left| \hat{W}_{j}^{2} \right| + VAR[s^{(k)}(q)]. \quad (3-8)$$

Hence, the output accuracy $A_{q}$ at the q-th output, in terms of the number of fractional bits, is approximately given by:
will be optimized together with the other vectors precisely, we can formulate the problem as follows: the intermediate signal formats. The one with the minimum wordlengths which are found by the random search algorithm. This can ensure that no overflow will occur at any adder output, at the expense of slightly increased hardware complexity. In the next subsection, we will describe the approach to determine the internal wordlength with prescribed output accuracy.

C. — Wordlength determination

For a given output accuracy, the idea of the random search is explored to minimize the hardware complexity [8,9] using the proposed noise model of the FFT algorithm. To start with, an objective function regarding the hardware complexity has to be set up. As an illustration, the number of adder cells and/or registers is employed as a measure of the hardware complexity since it is always the major resources in the hardware environment. Also, it is related to the internal wordlengths for the intermediate signals which are the variables that we want to optimize. Note that other meaningful measures can also be used instead. In general, the determination of the internal wordlength can be done in three steps. First of all, the SOPOT approximation of the twiddle factors is found as discussed in section II-B. Secondly, the format of maximum wordlengths for the intermediate signals, including both real and imaginary parts can be obtained by assuming all the inputs are in their maximum values and the precisions of all the signals are retained. Finally, from the sections III-A and III-B, the noise powers introduced by the rounding operation as well as the output bit accuracies can be statistically calculated, with respect to the proposed wordlengths which are found by the random search algorithm. These proposed wordlengths are stored in the vector \( \delta \) which will be optimized together with the other vectors \( \delta \) storing all the intermediate signal formats. The one with the minimum number of adder cells, while satisfying the prescribed output accuracy, will be declared as the solution of the problem. More precisely, we can formulate the problem as follows:

\[
\min_{(\delta, \delta)} C(\delta, \delta_j) \text{ subject to } P_{\text{total}} \leq P_{\text{spec}},
\]

where \( P_{\text{total}} \) and \( P_{\text{spec}} \) are respectively the total noise power and the specified output accuracy at the DFT output, and \( C(\cdot) \) is the objective function in this problem. There are two ways to speed up the search process. The first one is to identify the symmetries stage by stage because of the fact that all the input signals are assumed to be the maximum so that the number of variables can be largely reduced. For instance, to calculate the wordlengths for all the intermediate signals within a radix-2 64-point FFT, only the first two output points at the first stage are required to examine, rather than all 64 output points. The second one is to make a reasonable initial guess in order to shorten the search time. Higher noise power, say, is allowed at the earlier stage. Thus, shorter wordlength is allocated at the earlier stage such that the overall output bit accuracies at the final stage still satisfy the requirement.

IV. Design Example

This example shows the effectiveness of the round-off noise model as proposed in the section III. To start with, let’s consider the 64-point radix-2 (i.e. \( p = 2 \)) FFT with the prescribed accuracy at each output equal to 16. As mentioned earlier, the 2-point DFT can be implemented without any multiplications and its outputs at each stage are scaled by a factor of 1/2. The input sequence \( \{x(n)\}, n = 0, 1, \ldots, 63 \) are complex values with both real and imaginary parts having the format of \(<1/13>\) i.e. 14 bits with the maximum value equal to 0.99988. After the wordlength optimization as discussed in section III, the entire FFT structure requires 36070 adder cells. Table 1 and 2 show a summary of results and internal wordlength formats at the 6-th stage of the ML-FFT. The output formats at the remaining stages are omitted due to page limitation. To give an idea of the hardware savings of the proposed structure, a comparison with the structure using fixed wordlength is considered below. For the sake of the comparison, the internal wordlengths for all intermediate signals are fixed to 19 bits. The corresponding number of adder cells is 36936 which is slightly higher than that of the proposed one. For simulation purpose, we use Matlab to model the hardware implementation of the ML-FFT and assume that both structures are free from the coefficient round-off errors. Figure 4 shows the output bit accuracies of the radix-2 64-point ML-FFT using proposed wordlength (solid line) and fixed wordlength (dotted line), taking an average over 10000 random generated binary data with the format of \(<1/13>\). Result shows that our proposed structure meets the required bit accuracy quite well with slight deviation upon the 16-bit accuracy. Also, it shows that the output accuracies of the proposed structure are in general higher than that of the structure using fixed wordlength, except for those at output point 0, 15, 31 and 47. The reason is trivial due to the fact that there is no non-trivial multiplication when tracing the data path associated with those 4 output points from the first stage to the final stage. On the other hand, figure 4 also reveals that the more the multiplications with the twiddle factors, the lower the bit accuracies or the higher the round-off errors for that output point are suffered. As we expected, our proposed approach can efficiently control round-off errors by adjusting the internal wordlengths so that the prescribed output bit accuracies are satisfied without any overflow.

V. Conclusion

An error analysis of ML-FFT, using the DIT radix-\( p \) FFT as an example with \( N \) being an integer power of \( p \), is presented. ML-FFT parameterizes the twiddle factors in the conventional radix-\( p \) FFT algorithm as certain rotation-like matrices and approximates the associated parameters by sum-of-power-of-two (SOPOT) or canonical signed digits (CSD) representations. Apart from the error due to the SOPOT approximation, there is another error called signal round-off error which also affects the output bit accuracy of the ML-FFT. A recursive noise model is developed to model the statistics properties of these errors. By using this model, a random search algorithm is proposed to efficiently determine the minimum hardware complexity to realize the ML-FFT subject to the prescribed output bit accuracy. Simulation results show good agreement with the theoretical results.
REFERENCES


Figure 1: Block diagram of the DIT radix-p N-point FFT, N = p^n.

Figure 2: Proposed noise model for the DIT radix-p N-point FFT, N = p^n.

Figure 3: Implementation of the rotation-like matrix (refer to (2-4)) with round-off noises r_i for i = 0, 1, 2.

Table 1: Proposed wordlengths of the output formats at the 6-th stage (TF = Twiddle Factor, Real. = Real Storage and Imag. = Imaginary Storage).

Table 2: Proposed wordlengths of the output formats at the 6-th stage for the 64-point radix-2 ML-FFT (O/P = Output Point, Real. = Real Storage and Imag. = Imaginary Storage).