Digital Signal Processing 25 (2014) 280–288

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Efficient algorithms for computing the new Mersenne number transform

Mounir T. Hamood [a,*], Said Boussakta [b]

[a] Department of Electrical Engineering, University of Tikrit, Tikrit, PO Box 42, Iraq
[b] School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

ARTICLE INFO

Article history:
Available online 5 November 2013

Keywords:
Number theoretic transforms (NTTs)
New Mersenne number transform (NMNT)
Radix-2^2 algorithm

ABSTRACT

The new Mersenne number transform (NMNT) has proved to be an important number theoretic transform (NTT) used for error-free calculation of convolutions and correlations. Its main feature is that for a suitable Mersenne prime number (p), the allowed power-of-two transform lengths can be very large. In this paper, efficient radix-2^2 decimation-in-time and in-frequency algorithms for fast calculation of the NMNT are developed by deriving the appropriate mathematical relations in finite field and applying principles of the twiddle factor unscrambling technique. The proposed algorithms achieve both the regularity of radix-2 algorithm and the efficiency of radix-4 algorithm and can be applied to any powers of two transform lengths with simple bit reversing for ordering the output sequence. Consequently, the proposed algorithms possess the desirable properties such as simplicity and in-place computation. The validity of the proposed algorithms has been verified through examples involving large integer multiplication and digital filtering applications, using both the NMNT and the developed algorithms.

1. Introduction

Convolutions and correlations are the most fundamental mathematical tools used for enormous area of digital signal/image processing and other diverse applications [1,2]. For instance, convolutions are widely used in the design and implementation of the finite impulse response (FIR) as well as the infinite impulse response (IIR) digital filters. Moreover, it is well known that the DFT of prime lengths can be computed by converting it to a cyclic convolution using 'Rader's convolution algorithm' [3]. Correlation differs from convolution only by a simple inversion of one of the input sequences [4], therefore developments for the convolutions algorithms are equally applicable to the correlation also.

By proper scaling of the convolution's inputs, they can be always converted to a set of integers, and the convolution can be performed modulo a prime number M in the finite (Galois) field $GF(M)$. If the scaling factor is such that the convolution output has never exceeded $M/2$, then the convolution output has the identical values modulo M that would be obtained in the normal field. Under these conditions, the calculation of the convolution can be simplified by introducing a new family of transforms defined in finite field, known as number theoretic transforms (NTTs) [5,6], that have the same structure as the DFT but with complex operations replaced by an exact integer operations performed modulo M. NTTs first presented by Pollard [7], are discrete transforms defined over residue class fields or rings of integers, which were introduced for efficient calculation of error-free convolution and correlation without truncation or round-off errors.

NTTs have been firmly recognised within the field of signal processing [2]. Interesting applications of NTTs are found in the areas of digital filtering, image processing [8,9], fast coding and decoding [10], large integer and matrix multiplication [11,12], cryptography [13], and deconvolution [14]. This is owing to their contributing ability to perform error-free calculations over a field or a ring of integers whilst maintaining the cyclic convolution property (CCP). This is in contrast to other methods of calculation, such as the DFT which involves complex arithmetic with rounding and/or truncation errors in its calculations; errors also arise in the multiplication with cosine and sine functions which are irrational, preventing exact representation in a finite precision machine [15].

The most recognised NTTs are the Fermat (FNT) [16] and Mersenne (MNT) [6] number transforms. However, for standard signal processing applications the main drawback of these transforms is the stringent relationship between word length (the number of bits in the modulus), obtainable transform length, and a limited choice of possible word lengths. To retain the advantages of NTTs, the New Mersenne Number Transform (NMNT) was introduced [17,18], which alleviate this relationship. The NMNT is defined modulo the Mersenne numbers, where arithmetic operations are simple equivalent to 1's complement and has the cyclic convolution property; hence, it can be used for fast calculation of error-free convolutions and correlations. The NMNT is a

* Corresponding author.
E-mail addresses: m.t.hamood@tu.edu.iq (M.T. Hamood), said.boussakta@ncl.ac.uk (S. Boussakta).

1051-2004/$ – see front matter
http://dx.doi.org/10.1016/j.dsp.2013.10.018

particularly interesting NTT as it has a long powers of two lengths up to 2^p, making it amenable to fast algorithms.

Various Cooley–Tukey algorithms for the fast calculations of the NMNT have been developed based on both DIT and DIF approaches such as radix-2 [17,18], radix-4 [19,20] and split-radix [21,22] algorithms. However, for any transform to stand as a good candidate for real applications, its complete fast algorithms need to be developed.

Over the last years, a new hardware-oriented FFT algorithm known as radix-2^2 [23–25], as well as its variants algorithms [26–29], has been recognised as one of the most powerful structures used in pipeline architectures. It achieves at the same time both a simple and regular butterfly structure as radix-2 algorithm and a reduced number of twiddle factor multiplication provided by radix-4 algorithm. Therefore, it is desirable to generalise this algorithm to other discrete transforms such as the NMNT.

Therefore, the aim of this paper is to introduce new radix-2^2 decimation-in-time (DIT) and in-frequency (DIF) NMNT algorithms. The derivation of the proposed algorithms is based on the principle of the twiddle factor unscrambling technique [30–32], which is different from the conventional multidimensional index mapping technique [18]. The development of the presented algorithms has rested mainly on the observation that a radix-4 algorithm can be modified so that the output is in bit-reversed order; if a normal radix-4 butterfly is used, the output is in base-4 reversed order. However, if the outputs of the four short length butterflies are modified to have their outputs in bit-reversed order, the output of the total radix-4 algorithm will be in bit-reversed order and not base-4 reversed order.

The remaining contents of this paper are organised as follows: Section 2 reviews the NMNT and its cyclic convolution property. In Sections 3 and 4, we propose radix-2^2 DIT- and DIF-NMNT algorithms, respectively. In Section 5, we study the performance of the proposed algorithms by analysing their arithmetic complexity and comparing them with existing NMNT algorithms. Section 6 introduces two examples for the presented algorithms. A conclusion is then given in Section 7.

2. The new Mersenne number transform

2.1. Transform definition

Let p be a prime and $Mp = 2^p - 1$ Mersenne numbers, which are primes for $p = 2, 3, 5, 7, 13, 17, 19, \ldots$, etc. The NMNT of an integer sequence $x(n)$ of length N is given by [17,18]:

$$X(k) = \left\langle \sum_{n=0}^{N-1} x(n)\beta(nk) \right\rangle_{Mp}, \quad k = 0, 1, \ldots, N-1 \tag{1}$$

and its inverse has exactly the same form:

$$x(n) = \left\langle N^{-1} \sum_{k=0}^{N-1} X(k)\beta(nk) \right\rangle_{Mp}, \quad n = 0, 1, \ldots, N-1 \tag{2}$$

where:

$$\beta(nk) = \beta_1(nk) + \beta_2(nk) \tag{3}$$

$$\beta_1(nk) = \left\langle \mathrm{Re}(\alpha_1 + j\alpha_2)^{nk} \right\rangle_{Mp} \tag{4}$$

$$\beta_2(nk) = \left\langle \mathrm{Im}(\alpha_1 + j\alpha_2)^{nk} \right\rangle_{Mp} \tag{5}$$

Also:

$$\alpha_1 = \pm\left\langle 2^q \right\rangle_{Mp}; \qquad \alpha_2 = \pm\left\langle -3^q \right\rangle_{Mp}; \quad q = 2^{p-2} \tag{6}$$

$\langle\rangle_{M_p}$ represents modulo Mp.

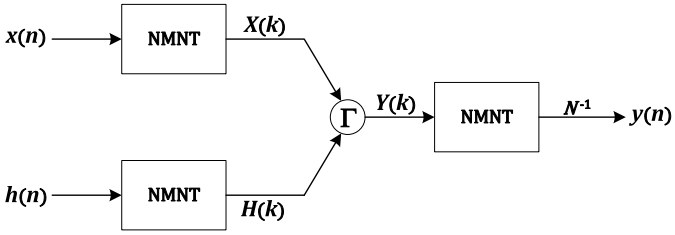

Fig. 1. Fast convolution using the NMNT.

α_1 and α_2 are of order $N = 2^{p+1}$. For transform length N/d where d is an integer power of two, β_1 and β_2 are given by:

$$\beta_1(nk) = \left\langle \mathrm{Re}\left((\alpha_1 + j\alpha_2)^d\right)^{nk} \right\rangle_{Mp} \tag{7}$$

$$\beta_2(nk) = \left\langle \mathrm{Im}\left((\alpha_1 + j\alpha_2)^d\right)^{nk} \right\rangle_{Mp} \tag{8}$$

Re(.) and Im(.) denote real and imaginary parts of the enclosed term respectively, (N^{-1}) exists and is given by (2^{p-d}), where $N = 2^d$ and d is an integer, $0 \leqslant d \leqslant p$.

2.2. NMNT cyclic convolution property

The NMNT has the cyclic convolution property; if $x(n)$ and $h(n)$ are two sequences to be convolved and $[y(n) = x(n) \circledast h(n)]$ is the convolution result, then

$$Y(k) = X(k)\boldsymbol{\Gamma} H(k) = X(k) \bullet H_{ev}(k) + X(N-k) \bullet H_{od}(k) \tag{9}$$

where $\circledast$ is the cyclic convolution operator and $\bullet$ is point-by-point multiplication. $X(k)$, $H(k)$ and $Y(k)$ stand for the NMNT transforms of $x(n)$, $h(n)$ and $y(n)$ respectively. $H_{ev}(k)$ and $H_{od}(k)$ stand for even and odd parts of $H(k)$ respectively, which are given by:

$$H_{ev}(k) = \left\langle \left(H(k) + H(N-k)\right) \times 2^{p-1} \right\rangle_{Mp} \tag{10}$$

$$H_{od}(k) = \left\langle \left(H(k) - H(N-k)\right) \times 2^{p-1} \right\rangle_{Mp} \tag{11}$$

If both $x(n)$ and $h(n)$ are properly padded with zeros, their circular convolution given in (9) will be equivalent to their linear convolution. To avoid overflow, the modulus, Mp must be chosen so that $y(n)$ does not exceed Mp, one upper bound is given by [5,18]:

$$\left|y(n)\right| \leqslant \left|x(n)\right|_{\max} \sum_{n=0}^{N-1} \left|h(n)\right| \leqslant Mp/2 \tag{12}$$

The process of calculation of the convolution via the NMNT is shown in Fig. 1, where the operator $\boldsymbol{\Gamma}$ is given in (9).

3. Decimation-in-time algorithm

The development of radix-2^2 algorithms starts by decomposing (1) into four partial sums and replacing (n) with $(4n + l)$ for $n = 0, 1, \ldots, N/4 - 1$ and $l = 0, 1, 2, 3$ as follows:

$$X(k) = \left\langle \sum_{l=0}^{3} \sum_{n=0}^{\frac{N}{4}-1} x(4n+l)\beta\left((4n+l)k\right) \right\rangle_{Mp} \tag{13}$$

According to (13), the input sequence $x(n)$ is decimated into four sets so that each partial sum represents NMNT of size $N/4$. The output sequence $X(k)$ is computed as four separate parts, and each part denoted by $X(k + \lambda N/4)$ has $(N/4)$ consecutive elements indexed by k for $k = 0, 1, \ldots, N/4 - 1$ and $\lambda = 0, 1, 2, 3$. Therefore, (13) becomes: