Unusual-length number-theoretic transforms using recursive extensions of Rader's algorithm

M.G. Parker
M. Benaissa

Indexing term: Number-theoretic transforms

Abstract: A novel decomposition of NTT blocklengths is proposed using repeated applications of Rader's algorithm to reduce the problem to that of realising a single small-length NTT. An efficient implementation of this small-length NTT is achieved by an initial basis conversion of the data, so that the new basis corresponds to the kernel of the small-length NTT. Multiplication by powers of the kernel become rotations and all arithmetic is efficiently performed within the new basis. More generally, this extension of Rader's algorithm is suitable for NTT or DFT applications where an efficient implementation of a particular small-length NTT/DFT module exists.

1 Introduction

The number-theoretic transform (NTT) has been suggested as an alternative to the DFT for computing cyclic convolution [1, 2] and is suitable for inclusion within signal processing, error-correction and residue number systems. Efficient architectures are possible for Fermat and Mersenne transforms [2–4], where multiplication within the transform is eliminated due to the correspondence of the NTT kernel to the basis of the data (i.e. the kernel is some simple power of 2 and the data are represented using a binary basis). However, there are only a few blocklengths over which these NTTs are possible. Various well known blocklength decomposition schemes can be used to widen the choice of NTT blocklengths [5–9] and, in this paper, a novel decomposition is suggested, based on successive applications of Rader's algorithm [10]. The technique is suitable for a wide range of prime and composite blocklengths and only requires repeated applications of a single, small-length NTT. It is, therefore, highly suitable for reduced hardware systems. Furthermore, by applying a preliminary basis conversion and matching the basis of the data to the kernel of the small-length NTT [11–13], all kernel multiplications can be eliminated, and the only multiplications are the fixed multiplications, inherent in the NTT-based cyclic convolutions, which are required to realise Rader's algorithm. Although this paper develops the theory in terms of the NTT, the extension of Rader's algorithm is equally valid for DFTs, and may be combined with efficient Winograd solutions for small-length DFTs [7] to construct unusual-length DFTs.

2 Theory

The N-point NTT is given by

$$X[k] = \left\langle \sum_{n=0}^{N-1} x[n]\alpha^{nk} \right\rangle_M \quad 0 \leqslant k < N \tag{1}$$

where α is an Nth root of 1, mod M and, for the purposes of this discussion, M is considered prime.

Consider the case where

$$N_p = N + 1 \quad N_p \text{ prime} \tag{2}$$

Then a one-dimensional NN_p-point NTT can always be defined, mod M, where

$$M = rNN_p + 1 \tag{3}$$

for r integer, positive, and M prime. This NN_p-point NTT can be expressed as

$$X[k] = \left\langle \sum_{n=0}^{NN_p-1} x[n](\alpha\beta^{-1})^{nk} \right\rangle_M \quad 0 \leqslant k < NN_p \tag{4}$$

where α, β and $(\alpha\beta^{-1})$ are Nth, N_pth and NN_pth roots of 1, mod M, respectively.

As gcd $(N, N_p) = 1$, eqn. 4 can be reformulated using the prime factor algorithm (PFA) [2] as a two-dimensional NTT,

$$X[k_0, k_1] = \left\langle \sum_{n_1=0}^{N_p-1} x[k_0, n_1]\beta^{n_1k_1} \right\rangle_M \quad 0 \leqslant k_1 < N_p \tag{5}$$

where

$$x[k_0, n_1] = \left\langle \sum_{n_0=0}^{N-1} x[n_0, n_1]\alpha^{n_0k_0} \right\rangle_M \quad 0 \leqslant k_0 < N \tag{6}$$

with $n = \langle N_p\langle N_p^{-1}\rangle_N, n_0 + N\langle N^{-1}\rangle_{N_p}n_1\rangle_{NN_p}$ and $k = \langle N_pk_0 + Nk_1\rangle_{NN_p}$. Eqns. 5 and 6 comprise NN_p-point NTTs, mod M, and N_pN-point NTTs, mod M, respectively. If an efficient architecture exists for the computation of the N-point NTT, then the first dimension (eqn. 6) of the above 2D NTT is easily computed. In particular, the input data, $x[n]$, can, for a given n, be preconverted to an α-basis, such that

$$x(n) = \left\langle \sum_{i=0}^{l-1} x_i\alpha^i \right\rangle_M$$

$$\text{for } x_i \in \{S\} \text{ and } \alpha \text{ an } N\text{th root of 1, mod } M \tag{7}$$

where S is a small set of integers (in the extreme, $S = \{0, 1\}$) and l is equal to or slightly less than some

Paper 1689K (E5), first received 31st May and in revised form 31st August 1994

The authors are with the School of Engineering, University of Huddersfield, Huddersfield HD1 3DH, United Kingdom

multiple of N. In such a case, the NTT of eqn. 6 can be computed without multiplication, the multiplications by powers of α being replaced with rotations of $x[n]$ [11, 12]. (These principles are simply generalisations of the ideas inherent within the implementations of Fermat or Mersenne number theoretic transforms [2–4, 14]). To compute the N_p-point NTTs, eqn. 5, multiplication by powers of the kernel, β, are required. These NTTs will not be simply realised in an α-basis as β cannot be a simple power of α. It is possible to convert the data from an α-basis to a β-basis after having performed the N-point NTTs and prior to performing the N_p-point NTTs [13]. However, suitable basis converters may not always be feasible and, for word-parallel implementations, a large number of basis converters will be required. An alternative solution, the subject of this paper is to utilise Rader's algorithm [10] to compute the N_p-point NTTs using N-point NTTs, mod M.

Rader's algorithm states that a P-point DFT, where P is prime, can be computed using a $P - 1$-point complex cyclic convolution (CC). This $P - 1$-point CC can, in turn, be computed using any orthogonal transform (such as the NTT), possessing the 'cyclic convolution property'. Similarly, a P-point NTT, mod M, can be computed using a $P - 1$-point CC, mod M, which can, in turn, be computed using $P - 1$-point NTT/INTTs, mod M, *providing* a $P - 1$-point NTT exists, mod M. From eqn. 3 it is clear that M supports N_p *and* N-point NTTs. Therefore the N_p-point NTT of eqn. 5 can be represented as

$$Y[k] = \left\langle \sum_{n=0}^{N_p-1} y[n]\beta^{nk} \right\rangle_M$$

and decomposed as follows

$$Y[0] = \left\langle \sum_{n=0}^{N_p-1} y[n] \right\rangle_M \tag{8}$$

$$Y[k] = \langle Y'[q] + y[0] \rangle_M \quad 1 \leqslant k < N_p \tag{9}$$

where:

$$Y'[q] = \left\langle \sum_{p=0}^{N_p-2} y'[p]\beta'_{q-p} \right\rangle_M \quad 0 \leqslant q < N_p - 1 \tag{10}$$

$$n = \langle g^{-p} \rangle_{N_p} \quad k = \langle g^q \rangle_{N_p} \tag{11}$$

$$y'[p] = y[\langle g^{-p} \rangle_{N_p}] = y[n]$$

$$\beta'_{q-p} = \beta^{\langle g^{q-p} \rangle_{N_p}} = \beta^{nk} \tag{12}$$

$$Y'[q] = \langle Y[\langle g^q \rangle_{N_p}] - y[0] \rangle_M \tag{13}$$

and g is an $(N_p - 1)$th root of 1, mod N_p. Thus, using N-point NTT/INTTs to compute the $N = N_p - 1$-point CC of eqn. 10,

$$Y'[q] = \left\langle N^{-1} \sum_{k=0}^{N-1} F(k)G(k)\alpha^{-qk} \right\rangle_M \quad 0 \leqslant q < N \tag{14}$$

where

$$F(k) = \left\langle \sum_{p=0}^{N-1} y'[p]\alpha^{pk} \right\rangle_M \quad 0 \leqslant k < N \tag{15}$$

and

$$G(k) = \left\langle \sum_{p=0}^{N-1} \beta'_p \alpha^{pk} \right\rangle_M \quad 0 \leqslant k < N \tag{16}$$

Eqn. 16 can be precomputed, and $G(0)$ is always equal to -1. Furthermore, one other value of $G(k)$ will be a power of α. Therefore the CC of eqn. 10 can be computed in an α-basis using only $N - 2$ fixed (nontrivial) multiplications. This can be considered a great improvement on the direct α-basis implementation of an N_p-point NTT, mod M, which requires approximately $(N_p - 1)^2$ fixed multiplications. This PFA/Rader-based NN_p-point NTT can be compared to a direct, undecomposed version and a PFA-based version, as follows.

An undecomposed NN_p-point NTT, mod M, will require

$(NN_p - 1)^2$ fixed multiplications,
mod M (ignoring additions).

Using an $N \times N_p$ PFA decomposition, the NN_p-point NTT will require

$N_p(N - 1)^2 + N(N_p - 1)^2$ fixed multiplications,
mod M (ignoring additions)

The PFA/Rader decomposition of the NN_p-point NTT, as described in this paper, will require

$(N_p + 2N)$ N-point NTTs and $N(N - 2)$
fixed multiplications, mod M

Assuming that each N-point NTT is performed efficiently without multiplication in an α basis, then the complexity of each NTT can be approximately equated as

1 N-point NTT, mod $M \simeq N^2/\lceil \log_2(M) \rceil$
fixed multiplications, mod M

Therefore the PFA/Rader method requires approximately

$(N_p + 2N)N^2/\log_2(M) + N(N - 2)$
fixed multiplications, mod M

As a further approximation, if one assumes $N_p \simeq N$, the above estimates reduce to

Undecomposed NN_p-point NTT,
mod $M \simeq N^4$ fixed multiplications, mod M

PFA-based NN_p-point NTT,
mod $M \simeq 2N^3$ fixed multiplications, mod M

PFA/Rader-based NN_p-point NTT,
mod $M \simeq 3N^3/\lceil \log_2(M) \rceil + N^2$
fixed multiplications, mod M (17)

The figure for the PFA/Rader-based NTT compares favourably with the other methods. Furthermore, all additions and multiplications will be performed in an α-basis, and, as shown in [15], particularly efficient implementations of modular arithmetic operations are possible if the basis α satisfies eqn. 7 for a given modulus, M.

As an example, let $N = 16$, $N_p = 17$, $M = 1361$. Note that $\alpha = 63$, where 63 has order 16, mod 1361. All integers, mod 1361, can be represented using a 63-basis, as specified in eqn. 7, where l is a minimum of 16 and $S = \{0, 1\}$, i.e. all data, mod 1361, can be represented using 16-bit words, where each consecutive bit represents a consecutive power of 63. Furthermore all 16-point NTT kernel products, mod 1361, can be implemented as bit rotations. Therefore a $16 \times 17 = 272$-point NTT, mod 1361, can be implemented using $17 + 32 = 49$ multiplierless 16-point NTTs, mod 1361, and $16 \times 14 = 224$ fixed multiplications for the 16-point CCs, with all arithmetic performed in a 63-basis. Using the estimates of eqn. 17:

Undecomposed 272-point NTT,
mod 1361 $\simeq$ 65 536 fixed multiplications, mod 1361

PFA/Rader-based 272-point NTT,
mod 1361 $\simeq$ 1374 fixed multiplications, mod 1361

PFA/Rader-based 272-point NTT,
mod 1361 $\simeq$ 1374 fixed multiplications, mod 1361

3 Recursive extensions of Rader's algorithm

Further Rader-based, composite NTTs can be constructed from the minimally composite NTTs of the preceding Section using prime-factor techniques. Thus, if

$$N_{p2} = NN_{p1} + 1 \quad \text{with } N_{p1}, N_{p2} \text{ prime} \qquad (18)$$

then N_{p2}-point NTTs can be computed, mod M, using NN_{p1}-point NTTs, as described in the preceding Section, where

$$M = r(NN_{p1}N_{p2}) + 1$$
$$\text{for } r \text{ integer, positive and } M \text{ prime} \qquad (19)$$

As N_{p2} is mutually prime to both N and N_{p1}, any composite NTT, mod M, which comprises any or all of N, N_{p1} and N_{p2}, can be computed using only N-point NTTs, mod M.

For example, if $N = 12$, $N_{p1} = 13$, then $N_{p2} = 157$ and 24492 ($= 12 \times 13 \times 157$), 1884, ($= 12 \times 157$) and 2041 ($= 13 \times 157$)-point NTTs can all be defined, mod M ($= r \times 24492 + 1$), using only 12-point NTT modules.

These sequences of N_{p_i} can be iterated as long as all N_{p_i} are prime. Unfortunately few of these 'Rader sequences' continue for more than one or two times before reaching a nonprime. Table 1 shows some example sequences for selected N. Note that, as a special case, a $2 \times 3 \times 7 \times 43 = 1806$-point NTT can be computed, mod M, where $M = r \times 1806 + 1$, using only 2-point NTTs, mod M.

Table 1: Rader sequences for selected *N*

N	Rader sequence
2	3, 7, 43
4	5
6	7. 43
10	11
12	13, 157
...	...
66	67, 4423, ..., etc.
...	...
192	193, 37057, ..., etc.
...	...
456	457, 208393, ..., etc.

For a given N, the sequence set can be extended to include any primes of the form $P + 1$, where P is generated as a multiplicative product of any or all members of the set $\{N, 2, N_{p1}, N_{p2}, \ldots\}$. (Note, that 2 is included as 2-point NTTs are simply implemented and never require multiplication.) Table 2 shows a selection of extended Radar sequences.

Although the sequences in Table 2 include a wide selection of primes, the most effective combinations will include as few 2-point and as many N-point NTTs as possible. A further broadening of the sequences (not shown in Table 2), is possible by also including all t, where $t \mid N$. This is justified because, if N-point NTTs are efficiently implemented, without multiplication, in an α-basis, then t-point NTTs will also be efficiently implemented without multiplication.

As a final example, a $5 \times 11 \times 23 = 1265$-point NTT, mod 245411, can be implemented using an $\alpha = 22$-basis, where 22 has order 5, mod 245411, $l = 5$ and $\boldsymbol{S} = \{0, 1, 2, \ldots, 21\}$. The 1265-point NTT uses only 5-point and 2-point NTTs, mod 245411, with all multiplications by powers of 22 implemented as rotations. Note, the blocklengths 11 and 23 are decomposed, using the PFA and Radar algorithms, as $(5 \times 2) + 1$ and $(((5 \times 2) + 1) \times 2) + 1$, respectively.

Table 2: Extended rader sequences for selected *N*

N	Extended Rader sequence
2	3, 7, 43
3	2, 7, 43
4	2, 5, 3, 11, 13, 7, 61, 31, 23, 67, 661, 331, 53, 131, 157, etc.
5	2, 3, 11, 7, 31, 23, 67, 331, 71, 43, 211, 463, 2311, 311, etc.
6	2, 7, 3, 43, etc.
7	2, 3, 43
8	2, 3, 7, 43, 1033, 24793, 6199, etc.
9	2, 3, 19, 7, 127, 43, 2287, 14479, etc.
10	2, 11, 3, 23, 31, 7, 331, 67, 47, 2531, 691, 139, 7591, 311, etc.
11	2, 3, 23, 7, 67, 47, 139, 43, 463, 967, 10627, 4423, 3083, etc.
12	2, 13, 3, 157, 7, 79, etc.
66	2, 67, 3, 4423, 7, 26539, 463, 43, etc.
192	2, 193, 3, 37057, 7, 43, 57793, 348559, etc.
456	2, 457, 3, 208393, 7, 43, 19609, etc.

The complete 1265-point NTT, mod 245411, requires

1593 5-point NTTs

4560 2-point NTTs

3180 fixed multiplications

and initial 2 to 22 basis conversions for all incoming data

Unlike the previous example, where $\alpha = 63$ and $\boldsymbol{S} = \{0, 1\}$ (requiring 1 bit), in this example the set $\boldsymbol{S}$ 'spans' the basis (and requires 5 bits). In other words, a representation for all integers $\{0, 1, \ldots, \alpha - 1\}$ is contained in $\boldsymbol{S}$. In this example, $\{0, 1, \ldots, 2\} \in \boldsymbol{S}$. This spanning guarantees a simple implementation for addition and, consequently, general multiplication, mod 245411, as additive carry propagation is localised [15]. Moreover, $\boldsymbol{S}$ can be widened up to $\{0, 1, 2, \ldots, 31\}$ (whilst still requiring 5 bits), and all computations can now use redundant arithmetic structures [16] with reduced carry propagations and, consequently, increased speed. The greatest drawback with this particular NTT example is the increased wordlength requirements, from $\lceil \log_2(245411) \rceil = 18$ bits, using a conventional binary representation, to $5 \times 5 = 25$ bits using a 22-basis with $\boldsymbol{S} = \{0, 1, 2, \ldots, 31\}$. It is hoped that more competitive Rader-based NTTs will become apparent over larger moduli.

4 Conclusion

A repeated application of Rader's algorithm has been proposed for the realisation of unusual-length NTTs. It allows the construction of relatively long-length NTTs using a single, small-length NTT. This small-length NTT can be efficiently implemented, without multiplication, by a preliminary basis conversion, so that the basis representation of the data corresponds to the kernel of the small-length NTT [15]. Thus the multiplication count of the complete NTT is reduced to the relatively small number of point-product multiplications, inherent within each application of Rader's algorithm. The method has been developed in conjunction with a prime-factor decomposition and, as the blocklengths comprise unusual primes, these NTTs are easily combined with more conventional NTTs (such as FNTs), using prime-factor techniques, to contruct even larger NTTs. This recursive Rader algorithm is particularly suited to reduced-hardware solutions and is also applicable to any DFT for which an efficient small-length DFT module exists.

5 References

1 POLLARD, J.M.: 'The fast Fourier transform in a finite field', *Math. Comput.*, 1971, **25**, (114), pp. 266–273

2 McCLELLAN, J.H., and RADER, C.M.: 'Number theory in digital signal processing' (Prentice-Hall, 1979)

3 McCLELLAN, J.H.: 'Hardware realisation of a Fermat number transform', *IEEE Trans.*, 1976, **ASSP-24**, (3), pp. 216–225

4 BENAISSA, M., DLAY, S.S., and HOLT, A.G.J.: 'CMOS VLSI design of a high-speed FNT based convolver/correlator using 3-input adders', *IEE Proc. G*, 1991, **138**, (2), pp. 182–190

5 GOOD, I.J.: 'The interaction algorithm and practical Fourier analysis', *J. Roy. Stat. Soc. B*, 1958, **20**, (2), pp. 361–372

6 COOLEY, J.W., and TUKEY, J.W.: 'An algorithm for the machine calculation of complex Fourier series', *Math. Comput.*, 1965, **19**, (2), pp. 297–301

7 WINOGRAD, S.: 'On computing the discrete Fourier transform', *Math. Comput.*, 1978, **32**, (141), pp. 175–1299

8 BLAHUT, R.E.: 'Fast algorithms for digital signal processing' (Addison-Wesley, Reading, MA, 1985)

9 JONES, K.J.: 'High-throughput, reduced hardware systolic solution to prime factor discrete Fourier transform algorithm', *IEE Proc. E*, 1990, **137**, (3), pp. 191–196

10 RADER, C.M.: 'Discrete Fourier transforms when the number of data samples is prime', *Proc. IEEE*, 1968, **56**, pp. 1107–1108

11 PARKER, M.G., and BENAISSA, M.: 'A bit-serial, VLSI implementation of a 60-point NTT using binary and ternary bases'. Digest of Int. Symp. on *DSP for communication systems*, University of Warwick, 7–9 September 1992

12 PARKER, M.G., and BENAISSA, M.: 'Bit-serial, VLSI architecture for the implementation of maximum-length number-theoretic transforms using mixed basis representations'. Proceedings of ICASSP '93, Minneapolis, April 1993, Vol. I, pp. 341–344

13 PARKER, M.G., and BENAISSA, M.: 'VLSI structures for bit serial modular multiplication using basis conversion', *IEE Proc. E*, 1994, **141**, pp. 381–390

14 HONDA, M., KAMEYAMA, M., and HIGUCHI, T.: 'Residue arithmetic based multiple-valued VLSI image processor'. Proceedings of the 22nd Int. Symp. on *Multivalued logic*, IEEE 1992, pp. 330–336

15 PARKER, M.G., and BENAISSA, M.: 'Using redundant number representations for efficient VLSI implementation of modular arithmetic'. Proceedings of IEE Colloquium on 'synthesis and optimisation of logic systems', (E3, E10), Savoy Place, London, 14 March 1994, Digest No. 1994/066

16 KNOWLES, S.C., and McWHIRTER, J.G.: 'The application of redundant number systems to the design of VLSI recursive filters'. Proceedings of IMA Conf. on *Mathematics in signal processing*, University of Warwick, 13–15 December 1988, pp. 27–42

17 TRUONG, T.K., REED, I.S., HSU, I.S., SHYU, H.C., and SHAO, H.M.: 'A pipeline design of a fast prime factor DFT on a finite field', *IEEE Trans.*, 1988, **C-37**, (3), pp. 365–374