$$k = (-1)^{m_1} 5^{m_2}$$
$$k^* = (-1)^{m_1^*} 5^{m_2^*}$$

and the reindexed core matrix gives rise to a two-dimensional convolution. The latter may be carried out by means of two 2D DFTs on $2 \times (N/4)$ points.

1.3.3.2.4. *The Winograd algorithms*

The cyclic convolutions generated by Rader's multiplicative reindexing may be evaluated more economically than through DFTs if they are re-examined within a new algebraic setting, namely the theory of congruence classes of polynomials [see, for instance, Blahut (1985), Chapter 2; Schroeder (1986), Chapter 24].

The set, denoted $\mathbb{K}[X]$, of polynomials in one variable with coefficients in a given field $\mathbb{K}$ has many of the formal properties of the set $\mathbb{Z}$ of rational integers: it is a *ring* with no zero divisors and has a *Euclidean algorithm* on which a theory of divisibility can be built.

Given a polynomial $P(z)$, then for every $W(z)$ there exist unique polynomials $Q(z)$ and $R(z)$ such that

$$W(z) = P(z)Q(z) + R(z)$$

and

$$\text{degree } (R) \; < \; \text{degree } (P).$$

$R(z)$ is called the *residue* of $H(z)$ modulo $P(z)$. Two polynomials $H_1(z)$ and $H_2(z)$ having the same residue modulo $P(z)$ are said to be *congruent* modulo $P(z)$, which is denoted by

$$H_1(z) \equiv H_2(z) \bmod P(z).$$

If $H(z) \equiv 0 \bmod P(z)$, $H(z)$ is said to be *divisible* by $P(z)$. If $H(z)$ only has divisors of degree zero in $\mathbb{K}[X]$, it is said to be *irreducible over* $\mathbb{K}$ (this notion depends on $\mathbb{K}$). Irreducible polynomials play in $\mathbb{K}[X]$ a role analogous to that of prime numbers in $\mathbb{Z}$, and any polynomial over $\mathbb{K}$ has an essentially unique factorization as a product of irreducible polynomials.

There exists a *Chinese remainder theorem* (CRT) for polynomials. Let $P(z) = P_1(z) \ldots P_d(z)$ be factored into a product of pairwise coprime polynomials [*i.e.* $P_i(z)$ and $P_j(z)$ have no common factor for $i \neq j$]. Then the system of congruence equations

$$H(z) \equiv H_j(z) \bmod P_j(z), \quad j = 1, \ldots, d,$$

has a unique solution $H(z)$ modulo $P(z)$. This solution may be constructed by a procedure similar to that used for integers. Let

$$Q_j(z) = P(z)/P_j(z) = \prod_{i \neq j} P_i(z).$$

Then P_j and Q_j are coprime, and the Euclidean algorithm may be used to obtain polynomials $p_j(z)$ and $q_j(z)$ such that

$$p_j(z)P_j(z) + q_j(z)Q_j(z) = 1.$$

With $S_i(z) = q_i(z)Q_i(z)$, the polynomial

$$H(z) = \sum_{i=1}^{d} S_i(z)H_i(z) \bmod P(z)$$

is easily shown to be the desired solution.

As with integers, it can be shown that the 1:1 correspondence between $H(z)$ and $H_j(z)$ sends sums to sums and products to products, *i.e.* establishes a *ring isomorphism*:

$$\mathbb{K}[X] \bmod P \cong (\mathbb{K}[X] \bmod P_1) \times \ldots \times (\mathbb{K}[X] \bmod P_d).$$

These results will now be applied to the efficient calculation of cyclic convolutions. Let $\mathbf{U} = (u_0, u_1, \ldots, u_{N-1})$ and $\mathbf{V} = (v_0, v_1, \ldots, v_{N-1})$ be two vectors of length N, and let $\mathbf{W} = (w_0, w_1, \ldots, w_{N-1})$ be obtained by cyclic convolution of $\mathbf{U}$ and $\mathbf{V}$:

$$w_n = \sum_{m=0}^{N-1} u_m v_{n-m}, \quad n = 0, \ldots, N-1.$$

The very simple but crucial result is that this cyclic convolution may be carried out by *polynomial multiplication modulo* $(z^N - 1)$: if

$$U(z) = \sum_{l=0}^{N-1} u_l z^l$$
$$V(z) = \sum_{m=0}^{N-1} v_m z^m$$
$$W(z) = \sum_{n=0}^{N-1} w_n z^n$$

then the above relation is equivalent to

$$W(z) \equiv U(z)V(z) \bmod (z^N - 1).$$

Now the polynomial $z^N - 1$ can be *factored* over the field of rational numbers into irreducible factors called *cyclotomic polynomials*: if d is the number of divisors of N, including 1 and N, then

$$z^N - 1 = \prod_{i=1}^{d} P_i(z),$$

where the cyclotomics $P_i(z)$ are well known (Nussbaumer, 1981; Schroeder, 1986, Chapter 22). We may now invoke the CRT, and exploit the ring isomorphism it establishes to simplify the calculation of $W(z)$ from $U(z)$ and $V(z)$ as follows:

(i) compute the d residual polynomials

$$U_i(z) \equiv U(z) \bmod P_i(z), \quad i = 1, \ldots, d,$$
$$V_i(z) \equiv V(z) \bmod P_i(z), \quad i = 1, \ldots, d;$$

(ii) compute the d polynomial products

$$W_i(z) \equiv U_i(z)V_i(z) \bmod P_i(z), \quad i = 1, \ldots, d;$$

(iii) use the CRT reconstruction formula just proved to recover $W(z)$ from the $W_i(z)$:

$$W(z) \equiv \sum_{i=1}^{d} S_i(z)W_i(z) \bmod (z^N - 1).$$

When N is not too large, *i.e.* for 'short cyclic convolutions', the $P_i(z)$ are very simple, with coefficients 0 or ± 1, so that (i) only involves a small number of additions. Furthermore, special techniques have been developed to multiply general polynomials modulo cyclotomic polynomials, thus helping keep the number of multiplications in (ii) and (iii) to a minimum. As a result, cyclic convolutions can be calculated rapidly when N is sufficiently composite.

It will be recalled that Rader's multiplicative indexing often gives rise to cyclic convolutions of length $p - 1$ for p an odd prime. Since $p - 1$ is highly composite for all $p \leq 50$ other than 23 and 47, these cyclic convolutions can be performed more efficiently by the above procedure than by DFT.

These combined algorithms are due to Winograd (1977, 1978, 1980), and are known collectively as 'Winograd small FFT algorithms'. Winograd also showed that they can be thought of as bringing the DFT matrix $\mathbf{F}$ to the following 'normal form':

$$\mathbf{F} = \mathbf{CBA},$$

where

$\mathbf{A}$ is an integer matrix with entries 0, ± 1, defining the 'pre-additions',

B is a diagonal matrix of multiplications,

C is a matrix with entries $0, \pm 1, \pm i$, defining the 'post-additions'.

The elements on the diagonal of **B** can be shown to be either real or pure imaginary, by the same argument as in Section 1.3.3.2.3.1. Matrices **A** and **C** may be rectangular rather than square, so that intermediate results may require extra storage space.

1.3.3.3. *Multidimensional algorithms*

From an algorithmic point of view, the distinction between one-dimensional (1D) and multidimensional DFTs is somewhat blurred by the fact that some factoring techniques turn a 1D transform into a multidimensional one. The distinction made here, however, is a practical one and is based on the dimensionality of the indexing sets for data and results. This section will therefore be concerned with the problem of factoring the DFT when the *indexing sets* for the input data and output results are multidimensional.

1.3.3.3.1. *The method of successive one-dimensional transforms*

The DFT was defined in Section 1.3.2.7.4 in an n-dimensional setting and it was shown that when the decimation matrix **N** is diagonal, say $\mathbf{N} = \mathrm{diag}(N^{(1)}, N^{(2)}, \ldots, N^{(n)})$, then $\bar{F}(N)$ has a tensor product structure:

$$\bar{F}(\mathbf{N}) = \bar{F}(N^{(1)}) \otimes \bar{F}(N^{(2)}) \otimes \ldots \otimes \bar{F}(N^{(n)}).$$

This may be rewritten as follows:

$$\begin{aligned}\bar{F}(\mathbf{N}) = {} & [\bar{F}(N^{(1)}) \otimes I_{N^{(2)}} \otimes \ldots \otimes I_{N^{(n)}}] \\ & \times [I_{N^{(1)}} \otimes \bar{F}(N^{(2)}) \otimes \ldots \otimes I_{N^{(n)}}] \\ & \times \ldots \\ & \times [I_{N^{(1)}} \otimes I_{N^{(2)}} \otimes \ldots \otimes \bar{F}(N^{(n)}],\end{aligned}$$

where the I's are identity matrices and $\times$ denotes ordinary matrix multiplication. The matrix within each bracket represents a one-dimensional DFT along one of the n dimensions, the other dimensions being left untransformed. As these matrices commute, the order in which the successive 1D DFTs are performed is immaterial.

This is the most straightforward method for building an n-dimensional algorithm from existing 1D algorithms. It is known in crystallography under the name of 'Beevers–Lipson factorization' (Section 1.3.4.3.1), and in signal processing as the 'row–column method'.

1.3.3.3.2. *Multidimensional factorization*

Substantial reductions in the arithmetic cost, as well as gains in flexibility, can be obtained if the factoring of the DFT is carried out in several dimensions simultaneously. The presentation given here is a generalization of that of Mersereau & Speake (1981), using the abstract setting established independently by Auslander, Tolimieri & Winograd (1982).

Let us return to the general n-dimensional setting of Section 1.3.2.7.4, where the DFT was defined for an arbitrary decimation matrix **N** by the formulae (where $|\mathbf{N}|$ denotes $|\det \mathbf{N}|$):

$$F(\mathbf{N}): \quad X(\mathbf{k}) = \frac{1}{|\mathbf{N}|} \sum_{\mathbf{k}^*} X^*(\mathbf{k}^*) e[-\mathbf{k}^* \cdot (\mathbf{N}^{-1}\mathbf{k})]$$

$$\bar{F}(\mathbf{N}): \quad X^*(\mathbf{k}^*) = \sum_{\mathbf{k}} X(\mathbf{k}) e[\mathbf{k}^* \cdot (\mathbf{N}^{-1}\mathbf{k})]$$

with

$$\mathbf{k} \in \mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n, \quad \mathbf{k}^* \in \mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n.$$

1.3.3.3.2.1. *Multidimensional Cooley–Tukey factorization*

Let us now assume that this decimation can be factored into d successive decimations, *i.e.* that

$$\mathbf{N} = \mathbf{N}_1\mathbf{N}_2 \ldots \mathbf{N}_{d-1}\mathbf{N}_d$$

and hence

$$\mathbf{N}^T = \mathbf{N}_d^T\mathbf{N}_{d-1}^T \ldots \mathbf{N}_2^T\mathbf{N}_1^T.$$

Then the coset decomposition formulae corresponding to these successive decimations (Section 1.3.2.7.1) can be combined as follows:

$$\begin{aligned}\mathbb{Z}^n &= \bigcup_{\mathbf{k}_1} (\mathbf{k}_1 + \mathbf{N}_1\mathbb{Z}^n) \\ &= \bigcup_{\mathbf{k}_1} \left\{ \mathbf{k}_1 + \mathbf{N}_1 \left[\bigcup_{\mathbf{k}_2} (\mathbf{k}_2 + \mathbf{N}_2\mathbb{Z}^n) \right] \right\} \\ &= \ldots \\ &= \bigcup_{\mathbf{k}_1} \ldots \bigcup_{\mathbf{k}_d} (\mathbf{k}_1 + \mathbf{N}_1\mathbf{k}_2 + \ldots + \mathbf{N}_1\mathbf{N}_2 \times \ldots \times \mathbf{N}_{d-1}\mathbf{k}_d + \mathbf{N}\mathbb{Z}^n)\end{aligned}$$

with $\mathbf{k}_j \in \mathbb{Z}^n/\mathbf{N}_j\mathbb{Z}^n$. Therefore, any $\mathbf{k} \in \mathbb{Z}/\mathbf{N}\mathbb{Z}^n$ may be written uniquely as

$$\mathbf{k} = \mathbf{k}_1 + \mathbf{N}_1\mathbf{k}_2 + \ldots + \mathbf{N}_1\mathbf{N}_2 \times \ldots \times \mathbf{N}_{d-1}\mathbf{k}_d.$$

Similarly:

$$\begin{aligned}\mathbb{Z}^n &= \bigcup_{\mathbf{k}_d^*} (\mathbf{k}_d^* + \mathbf{N}_d^T\mathbb{Z}^n) \\ &= \ldots \\ &= \bigcup_{\mathbf{k}_d^*} \ldots \bigcup_{\mathbf{k}_1^*} (\mathbf{k}_d^* + \mathbf{N}_d^T\mathbf{k}_{d-1}^* + \ldots + \mathbf{N}_d^T \times \ldots \times \mathbf{N}_2^T\mathbf{k}_1^* \\ &\quad + \mathbf{N}^T\mathbb{Z}^n)\end{aligned}$$

so that any $\mathbf{k}^* \in \mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n$ may be written uniquely as

$$\mathbf{k}^* = \mathbf{k}_d^* + \mathbf{N}_d^T\mathbf{k}_{d-1}^* + \ldots + \mathbf{N}_d^T \times \ldots \times \mathbf{N}_2^T\mathbf{k}_1^*$$

with $\mathbf{k}_j^* \in \mathbb{Z}^n/\mathbf{N}_j^T\mathbb{Z}^n$. These decompositions are the vector analogues of the multi-radix number representation systems used in the Cooley–Tukey factorization.

We may then write the definition of $\bar{F}(\mathbf{N})$ with $d = 2$ factors as

$$\begin{aligned}X^*(\mathbf{k}_2^* + \mathbf{N}_2^T\mathbf{k}_1^*) = {} & \sum_{\mathbf{k}_1} \sum_{\mathbf{k}_2} X(\mathbf{k}_1 + \mathbf{N}_1\mathbf{k}_2) \\ & \times e[(\mathbf{k}_2^{*T} + \mathbf{k}_1^{*T}\mathbf{N}_2)\mathbf{N}_2^{-1}\mathbf{N}_1^{-1}(\mathbf{k}_1 + \mathbf{N}_1\mathbf{k}_2)].\end{aligned}$$

The argument of $e(-)$ may be expanded as

$$\mathbf{k}_2^* \cdot (\mathbf{N}^{-1}\mathbf{k}_1) + \mathbf{k}_1^* \cdot (\mathbf{N}_1^{-1}\mathbf{k}_1) + \mathbf{k}_2^* \cdot (\mathbf{N}_2^{-1}\mathbf{k}_2) + \mathbf{k}_1^* \cdot \mathbf{k}_2.$$

The first summand may be recognized as a twiddle factor, the second and third as the kernels of $\bar{F}(\mathbf{N}_1)$ and $\bar{F}(\mathbf{N}_2)$, respectively, while the fourth is an integer which may be dropped. We are thus led to a 'vector-radix' version of the Cooley–Tukey algorithm, in which the successive decimations may be introduced in all n dimensions simultaneously by general integer matrices. The computation may be decomposed into five stages analogous to those of the one-dimensional algorithm of Section 1.3.3.2.1:

(i) form the $|\mathbf{N}_1|$ vectors $\mathbf{Y}_{\mathbf{k}_1}$ of shape $\mathbf{N}_2$ by

$$Y_{\mathbf{k}_1}(\mathbf{k}_2) = X(\mathbf{k}_1 + \mathbf{N}_1\mathbf{k}_2), \quad \mathbf{k}_1 \in \mathbb{Z}^n/\mathbf{N}_1\mathbb{Z}^n, \quad \mathbf{k}_2 \in \mathbb{Z}^n/\mathbf{N}_2\mathbb{Z}^n;$$

references