prime can itself be factored by invoking some extra arithmetic structure present in $\mathbb{Z}/p\mathbb{Z}$.

1.3.3.2.3.1. *N an odd prime*

The ring $\mathbb{Z}/p\mathbb{Z} = \{0, 1, 2, \ldots, p-1\}$ has the property that its $p-1$ non-zero elements, called *units*, form a *multiplicative group* $U(p)$. In particular, all units $r \in U(p)$ have a unique multiplicative inverse in $\mathbb{Z}/p\mathbb{Z}$, *i.e.* a unit $s \in U(p)$ such that $rs \equiv 1 \bmod p$. This endows $\mathbb{Z}/p\mathbb{Z}$ with the structure of a *finite field*.

Furthermore, $U(p)$ is a *cyclic* group, *i.e.* consists of the successive powers $g^m \bmod p$ of a generator g called a *primitive root mod p* (such a g may not be unique, but it always exists). For instance, for $p = 7$, $U(7) = \{1, 2, 3, 4, 5, 6\}$ is generated by $g = 3$, whose successive powers mod 7 are:

$$g^0 = 1, \quad g^1 = 3, \quad g^2 = 2, \quad g^3 = 6, \quad g^4 = 4, \quad g^5 = 5$$

[see Apostol (1976), Chapter 10].

The basis of Rader's algorithm is to bring to light a hidden regularity in the matrix $F(p)$ by permuting the basis vectors $\mathbf{u}_k$ and $\mathbf{v}_{k^*}$ of $L(\mathbb{Z}/p\mathbb{Z})$ as follows:

$$\begin{aligned}
\mathbf{u}'_0 &= \mathbf{u}_0 \\
\mathbf{u}'_m &= \mathbf{u}_k \quad \text{with } k = g^m, \qquad m = 1, \ldots, p-1; \\
\mathbf{v}'_0 &= \mathbf{v}_0 \\
\mathbf{v}'_{m^*} &= \mathbf{v}_{k^*} \quad \text{with } k^* = g^{m^*}, \quad m^* = 1, \ldots, p-1;
\end{aligned}$$

where g is a primitive root mod p.

With respect to these new bases, the matrix representing $\bar{F}(p)$ will have the following elements:

$$\begin{aligned}
\text{element } (0,0) &= 1 \\
\text{element } (0, m+1) &= 1 \quad \text{for all } m = 0, \ldots p-2, \\
\text{element } (m^*+1, 0) &= 1 \quad \text{for all } m^* = 0, \ldots, p-2, \\
\text{element } (m^*+1, m+1) &= e\left(\frac{k^* k}{p}\right) \\
&= e(g^{(m^*+m)/p}) \\
&\qquad \text{for all } m^* = 0, \ldots, p-2.
\end{aligned}$$

Thus the 'core' $\bar{C}(p)$ of matrix $\bar{F}(p)$, of size $(p-1) \times (p-1)$, formed by the elements with two non-zero indices, has a so-called *skew-circulant* structure because element (m^*, m) depends only on $m^* + m$. Simplification may now occur because multiplication by $\bar{C}(p)$ is closely related to a *cyclic convolution*. Introducing the notation $C(m) = e(g^{m/p})$ we may write the relation $\mathbf{Y}^* = \bar{F}(p)\mathbf{Y}$ in the permuted bases as

$$\begin{aligned}
Y^*(0) &= \sum_k Y(k) \\
Y^*(m^*+1) &= Y(0) + \sum_{m=0}^{p-2} C(m^*+m) Y(m+1) \\
&= Y(0) + \sum_{m=0}^{p-2} C(m^*-m) Z(m) \\
&= Y(0) + (\mathbf{C} * \mathbf{Z})(m^*), \quad m^* = 0, \ldots, p-2,
\end{aligned}$$

where $\mathbf{Z}$ is defined by $Z(m) = Y(p-m-2)$, $m = 0, \ldots, p-2$.

Thus $\mathbf{Y}^*$ may be obtained by cyclic convolution of $\mathbf{C}$ and $\mathbf{Z}$, which may for instance be calculated by

$$\mathbf{C} * \mathbf{Z} = F(p-1)[\bar{F}(p-1)[\mathbf{C}] \times \bar{F}(p-1)[\mathbf{Z}]],$$

where $\times$ denotes the component-wise multiplication of vectors. Since p is odd, $p-1$ is always divisible by 2 and may even be highly composite. In that case, factoring $\bar{F}(p-1)$ by means of the Cooley–Tukey or Good methods leads to an algorithm of complexity $p \log p$ rather than p^2 for $\bar{F}(p)$. An added bonus is that, because $g^{(p-1)/2} = -1$, the elements of $\bar{F}(p-1)[\mathbf{C}]$ can be shown to be either purely real or purely imaginary, which halves the number of real multiplications involved.

1.3.3.2.3.2. *N a power of an odd prime*

This idea was extended by Winograd (1976, 1978) to the treatment of prime powers $N = p^\nu$, using the cyclic structure of the multiplicative group of units $U(p^\nu)$. The latter consists of all those elements of $\mathbb{Z}/p^\nu\mathbb{Z}$ which are not divisible by p, and thus has $q_\nu = p^{\nu-1}(p-1)$ elements. It is cyclic, and there exist primitive roots g modulo p^ν such that

$$U(p^\nu) = \{1, g, g^2, g^3, \ldots, g^{q_\nu - 1}\}.$$

The $p^{\nu-1}$ elements divisible by p, which are divisors of zero, have to be treated separately just as 0 had to be treated separately for $N = p$.

When $k^* \notin U(p^\nu)$, then $k^* = pk_1^*$ with $k_1^* \in \mathbb{Z}/p^{\nu-1}\mathbb{Z}$. The results $X^*(pk_1^*)$ are p-decimated, hence can be obtained *via* the $p^{\nu-1}$-point DFT of the $p^{\nu-1}$-periodized data $\mathbf{Y}$:

$$X^*(pk_1^*) = \bar{F}(p^{\nu-1})[\mathbf{Y}](k_1^*)$$

with

$$Y(k_1) = \sum_{k_2 \in \mathbb{Z}/p\mathbb{Z}} X(k_1 + p^{\nu-1}k_2).$$

When $k^* \in U(p^\nu)$, then we may write

$$X^*(k^*) = X_0^*(k^*) + X_1^*(k^*),$$

where $\mathbf{X}_0^*$ contains the contributions from $k \notin U(p^\nu)$ and $\mathbf{X}_1^*$ those from $k \in U(p^\nu)$. By a converse of the previous calculation, $\mathbf{X}_0^*$ arises from p-decimated data $\mathbf{Z}$, hence is the $p^{\nu-1}$-periodization of the $p^{\nu-1}$-point DFT of these data:

$$X_0^*(p^{\nu-1}k_1^* + k_2^*) = \bar{F}(p^{\nu-1})[\mathbf{Z}](k_2^*)$$

with

$$Z(k_2) = X(pk_2), \qquad k_2 \in \mathbb{Z}/p^{\nu-1}\mathbb{Z}$$

(the $p^{\nu-1}$-periodicity follows implicity from the fact that the transform on the right-hand side is independent of $k_1^* \in \mathbb{Z}/p\mathbb{Z}$).

Finally, the contribution X_1^* from all $k \in U(p^\nu)$ may be calculated by reindexing by the powers of a primitive root g modulo p^ν, *i.e.* by writing

$$X_1^*(g^{m^*}) = \sum_{m=0}^{q_\nu - 1} X(g^m) e(g^{(m^*+m)/p^\nu})$$

then carrying out the multiplication by the skew-circulant matrix core as a convolution.

Thus the DFT of size p^ν may be reduced to two DFTs of size $p^{\nu-1}$ (dealing, respectively, with p-decimated results and p-decimated data) and a convolution of size $q_\nu = p^{\nu-1}(p-1)$. The latter may be 'diagonalized' into a multiplication by purely real or purely imaginary numbers (because $g^{(q_\nu/2)} = -1$) by two DFTs, whose factoring in turn leads to DFTs of size $p^{\nu-1}$ and $p-1$. This method, applied recursively, allows the complete decomposition of the DFT on p^ν points into arbitrarily small DFTs.

1.3.3.2.3.3. *N a power of 2*

When $N = 2^\nu$, the same method can be applied, except for a slight modification in the calculation of $\mathbf{X}_1^*$. There is no primitive root modulo 2^ν for $\nu > 2$: the group $U(2^\nu)$ is the direct product of *two* cyclic groups, the first (of order 2) generated by -1, the second (of order $N/4$) generated by 3 or 5. One then uses a representation

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

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',

references