New Area-Time Lower Bounds for the Multidimensional DFT

Gianfranco Bilardi **Carlo Fantozzi**

Department of Information Engineering
University of Padova,
Padova PD 35131, Italy,
Email: {bilardi,fantozzi}@dei.unipd.it

Abstract

Area-time lower bounds are derived for the VLSI computation of the $(n_1 \times n_2 \times \ldots \times n_d)$-point multidimensional DFT (MDDFT) over different types of rings and different types of input/output protocols.

First, an $AT^2 = \Omega\left((N \log |R|)^2\right)$ bound is obtained for any finite ring R, where $N = \prod_{k=1}^{d} n_k$, for any word-local protocol. The bound was previously known for the special case when $R = \mathbb{Z}_M$, the ring of integers modulo M.

Second, an $AT^2 = \Omega\left((Nb)^2\right)$ word-local bound is derived when R is the complex field, the components of the input are fixed-point numbers of b bits, and the precision of the output components guarantees that the resulting approximate transform is injective. No area-time lower bound was previously known for the DFT over the complex field.

Third, an $AT^2 = \Omega\left((N \log |R|)^2\right)$ bound is derived when $R = \mathrm{GF}\left(p^m\right)$ is a finite field of polynomials of degree m with coefficients in $\mathbb{Z}_p$, for certain classes of non word-local protocols. This is the first area-time lower bound derived for the DFT with I/O protocols that are not word-local.

Keywords: area-time tradeoff, digital signal processing, information exchange, multidimensional DFT, VLSI computation theory.

1 Introduction

The Fourier Transform plays a central role in the analysis of linear systems, in signal processing, in ordinary and partial differential equations, in the uncertainty relations of quantum mechanics, in the theory of error correcting codes, and several other areas. Furthermore, for several important computations, such as various forms of convolutions, polynomial multiplication, and integer multiplication, the most efficient algorithms known today are based on the Discrete Fourier Transform (DFT). Naturally, the DFT and its computation have been investigated extensively.

As computational systems evolve toward higher levels of parallelism, their performance is increasingly dominated by data movement. This has motivated the investigation of various measures of communication complexity of computational problems and algorithms. One such measure is the *information exchange.* Let Π be a computational problem and $\mathcal{V}$ a set of the variables that encode the input and the output of Π: the information exchange $I_{\mathcal{V}}(m)$ is defined as the minimum number of bits that two processors must exchange in order to solve Π under the constraint that exactly m of the variables of $\mathcal{V}$ are assigned to one of the two processors (each input and output variable being assigned to exactly one processor).

The information exchange has interesting relations with the *area-time* complexity in the *VLSI model of computation* (Abelson & Andreae 1980, Thompson 1980, Brent & Kung 1981, Ullman 1984). Bilardi & Preparata (1986) showed that the area A and the time T of a VLSI circuit satisfy the relation $AT^2 = \Omega\left(|\mathcal{V}| I_{\mathcal{V}}(m)^2/m\right)$, under reasonable assumptions on the I/O protocol of the VLSI circuit. Typically, a judicious choice of $\mathcal{V}$ and m is required, based on some insight on the structure of the computational problem. In (Wu 1984, Hornick & Sarrafzadeh 1987), area and information exchange are related via the bound $A = \Omega\left(I_{\mathcal{V}}(m)\right)$, under appropriate assumptions.

The pioneering work of Thompson (1979) has established the first lower bound to the information exchange I of the one-dimensional DFT for signals with N components that take value in the ring $\mathbb{Z}_M$ of the integers modulo M. Specifically, Thompson showed that $I \geq \lfloor (1/4) N \log_2 N \rfloor$, when the output components are equally split between the two processors. Correspondingly, $AT^2 = \Omega\left(N^2 \log^2 N\right)$. The result assumes that all bits that encode the same input or output component must be assigned to the same processor. This restriction is known as the *word-local* assumption. Thompson's bound has been extended – and somewhat strengthened, by replacing a $\log N$ factor by a $\log M$ factor – by Bilardi et al. (1989) who showed that, for the $(n_1 \times n_2 \times \ldots \times n_d)$-point MDDFT (multidimensional DFT) over $\mathbb{Z}_M$, $I \geq ((N-1)/4) \log_2 M$, where $N = n_1 n_2 \cdots n_d$. Here, the input variables are equally split between the two processors, in a word-local way. Correspondingly, $AT^2 = \Omega\left(N^2 \log^2 M\right)$.

The above results rely on the property that any $N/2 \times N/2$ submatrix of the MDDFT has rank at least $N/4$. For a finite ring R, this guarantees that the range of such a submatrix contains at least $|R|^{N/4}$ distinct outputs. Unfortunately, the connection between rank and number of distinct outputs is not immediate in the case of finite precision, hence the mentioned results do not cover the complex field. This is a significant limitation, since the complex field is by far the most frequently utilized ring in DFT applications. Motivated by these considerations, in Section 3 we develop a new approach, based on counting arguments that combine the injectivity and the time-

Support for the authors was provided in part by MIUR of Italy under project AlgoDEEP, by the University of Padova under Projects CPDA099949 and STPD08JA32, and by the Autonomous Province of Trento under the AURORA Project.

modulation/frequency-shift property of the Fourier transform. This approach leads to two results.

- An $AT^2 = \Omega\left((N \log |R|)^2\right)$ bound for any finite ring R, for any word-local protocol.
- An $AT^2 = \Omega\left((Nb)^2\right)$ word-local bound for the complex field, when the components of the input are fixed-point numbers of b bits and the precision of the output components guarantees that the resulting approximate transform is injective.

The result for finite rings could actually be derived by adaptations of the approach of Bilardi et al. (1989), where only the case $R = \mathbb{Z}_M$ had been considered. Instead, the result for the complex field is novel and does not appear to yield to the earlier approach.

All VLSI circuits proposed thus far in the literature for the DFT and the MDDFT are word-local. Whether a better area-time performance could be achieved by non word-local solutions is an interesting question. Indeed, this is known to be the case for some computational problems, such as sorting n keys of k bits each, when k grows more than logarithmically with n (Bilardi & Preparata 1985, 1986, Cole & Siegel 1988). To date, the only result for the DFT without the word-local restriction is an area lower bound, $A = \Omega(N \log M)$, derived in (Duris et al. 1985) for the one-dimensional DFT over the ring $\mathbb{Z}_M$, assuming the standard binary representation with $\lceil \log_2 M \rceil$ bits for the ring elements.

In Section 4, we turn our attention to the removal of the word-local assumption, which makes the study of the information exchange considerably harder. To gain some appreciation of the issues involved, consider that, in the word-local case, the representation chosen for the elements of the ring has no impact on the information exchange. The situation is different in the general case. For example, if $R = \mathbb{Z}_{q_1 q_2}$, with the q_1 and q_2 relatively prime, each element of R can be represented by a pair of integers, according to the Chinese Reminder Theorem. Then, the DFT over $\mathbb{Z}_{q_1 q_2}$ factors into a DFT over $\mathbb{Z}_{q_1}$ and one over $\mathbb{Z}_{q_2}$, with no information exchange between the two. No such decomposition of the computation is possible if the elements of $\mathbb{Z}_{q_1 q_2}$ are represented as integers modulo $q_1 q_2$. This observation generalizes to the decomposition of an arbitrary finite ring as a direct product of local rings (Dubois & Venetsanopoulos 1980).

To begin the exploration of non word-local protocols we have chosen to consider the important special case where the ring is a finite field. All finite fields of size p^m are algebraically isomorphic to each other and typically denoted as $\mathrm{GF}\left(p^m\right)$. However, the DFT on isomorphic fields may well exhibit different information exchange, if word-locality is not assumed. Hence, more specifically, we shall consider those representations of $\mathrm{GF}\left(p^m\right)$ where elements are viewed as polynomials of degree m, specified via their coefficients which belong to $\mathrm{GF}(p) = \mathbb{Z}_p$. These polynomials are added and multiplied modulo a given polynomial of degree m, irreducible over $\mathrm{GF}(p)$. The DFT over $\mathrm{GF}\left(p^m\right)$ has important applications in coding theory, particularly in connection with BCH error correcting codes (see Blahut (1983), Chapters 7-9). A number of algorithms for its computation have been proposed (Preparata & Sarwate 1977, Wang & Zhu 1988, Fedorenko 2006).

In Section 4.1, we introduce *semi word-local* protocols, where the constraint that all bits must be handled by the same port is assumed for the output components of the transform (word-local output), but only for the individual coefficients of the input components. We obtain the following result.

- If $R = \mathrm{GF}\left(p^m\right)$ is a finite field of polynomials, then $AT^2 = \Omega\left((N \log |R|)^2\right)$, for any semi word-local protocol.

This result provides some insight on how information moves across different coefficients of different degrees when the DFT is computed. Also novel is the lower bound technique, which exploits the characterization of the DFT provided by the convolution theorem.

In Section 4.2, we consider the field $\mathrm{GF}\left(p^2\right)$, where each element is represented by a pair of coefficients in $GF(p)$. We consider the protocol that assigns to one processor the first coefficient of the pair, for each input and output component, and assigns the second coefficient to the other processor. (Thus, neither the input nor the output is word-local). Although somewhat specialized, this case exhibits a rich behavior. If, in each dimension, the size n_k of the transform divides $p-1$, then the information exchange is null ($I = 0$); else, it is linear in the input size ($I = \Omega(N \log p)$). The latter bound is derived by a technique exploiting properties of the DFT on input signals with symmetries that are related to the notion of *conjugate* element within a field.

To complete the paper roadmap, in Section 2 we review the necessary basic notions and in Section 5 we present concluding remarks. While this paper focuses on area-time lower bounds, extensive work has been carried out on VLSI algorithms and architectures for the DFT, often providing matching upper bounds. VLSI circuits have been proposed for both the one-dimensional (Thompson 1980, 1983, Bilardi & Sarrafzadeh 1987, Bilardi et al. 1989, Yeh 2002) and the multidimensional (Chowdary & Steenaart 1984, Gertner & Shamash 1987, Chakrabarti & JáJá 1991, Alnuweiri 1994, Marino & Swartzlander 1999) DFT.

2 Basic Notions

2.1 The MDDFT

In a general, abstract formulation, the Fourier transform operates on signals of the form $x : G \to R$, where G is a *group* and R is a *commutative ring* with *identity*. In this paper, R will be either finite or the complex field. As for G, we will consider an arbitrary finite *abelian* group. Due to a well-known isomorphism result (Lidl & Pilz 1998, Theorem 10.26), $G \cong Z \triangleq \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \ldots \times \mathbb{Z}_{n_d}$. The group operation is addition of d-tuples, with the k-th component taken modulo n_k. Hereafter, we will simply write Z instead of $\mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \ldots \times \mathbb{Z}_{n_d}$, when there is no ambiguity about $n_1, n_2, \ldots, n_d$. A *d-dimensional signal* is a function $x : Z \to R$.

An element $\omega \in R$ is a *primitive n-th root of unity* if $\omega^0, \omega^1, \ldots, \omega^{n-1}$ are all distinct and $\omega^n = 1$. If, for $k = 1, 2, \ldots, d$, n_k has a multiplicative inverse ν_k in R, and the ring R is endowed with a primitive n_k-th root of unity ω_k, then the linear transformation

$$\sum_{j_1=0}^{n_1-1} \cdots \sum_{j_d=0}^{n_d-1} x(j_1, \ldots, j_d)\, \omega_1^{-i_1 j_1} \ldots \omega_d^{-i_d j_d} \qquad (1)$$

is called an $(n_1 \times n_2 \times \ldots \times n_d)$-point MDDFT. We refer to $x(\mathbf{j})$, $\mathbf{j} \in Z$, as the signal and to $X(\mathbf{i})$, $\mathbf{i} \in Z$, as its transform. The size of the MDDFT is $N = \prod_{k=1}^{d} n_k$. Equation 1 can be inverted, and the signal $x(\mathbf{j})$ recovered, by the linear transformation

$$\nu_1 \ldots \nu_d \sum_{i_1=0}^{n_1-1} \cdots \sum_{i_d=0}^{n_d-1} X(i_1, \ldots, i_d)\, \omega_1^{i_1 j_1} \ldots \omega_d^{i_d j_d}.$$

If R is a finite ring, then the existence of the transform is equivalent to the condition that $\mathrm{LCM}\,(n_1,\ldots,n_d)$ divides a certain integer $O[R]$ associated with the ring (Dubois & Venetsanopoulos 1978). If $R = \mathbb{Z}_M$, then $O[R] = \mathrm{GCD}\,(g_1-1,\ldots,g_l-1)$, where $g_1,\ldots,g_l$ are the prime factors of M; if $R = \mathrm{GF}\,(p^m)$, then $O[R] = p^m-1$ (Pollard 1971, Bonneau 1973, Agarwal & Burrus 1975).

As it is well known, if R is the complex field, for every $n > 0$ the set of primitive n-th roots of unity is given by $\{\exp(2\pi\sqrt{-1}\ell/n) : \mathrm{GCD}\,(\ell,n) = 1\}$, then the transform exists for any choice of $n_1, n_2, \ldots, n_d$. However, we can only compute approximate versions of the transform and therefore we need to be specific about the precision of the computation. Let us call (a,b)-*format* a fixed-point representation of complex numbers where both the real and the imaginary part are specified by a sign, a bits to the left of the binary point, and b bits to the right of the binary point. We call $(a,b;a',b')$-*complex MDDFT* the approximate version of the MDDFT where: (i) the components of the input signal are given in an (a,b)-format, and (ii) the components of the output are an (a',b')-format truncation of the infinite-precision transform of the input. The next proposition gives sufficient conditions for the invertibility of the approximate transform.

Proposition 1 *The* $(a,b;a+\lceil\log_2 N\rceil+1,b+1)$-*complex MDDFT is an invertible transformation.*

Proof We recall that the Euclidean norms of a signal x and its (exact) transform X satisfy the relation $||X||^2 = N||x||^2$. If x is in (a,b)-format, then $||x||^2 \leq 2^{2a+1}N$, hence $||X||^2 \leq 2^{2a+1}N^2$, or $||X|| \leq 2^{a+1/2}N$. Thus, $a+\lceil\log_2 N\rceil+1$ bits to the left of the binary point are sufficient to prevent overflow in the representation of the output.

We now want to show that, if u and v are different N-point signals in (a,b)-format, then the $(a+\lceil\log_2 N\rceil+1,b+1)$-format truncation of their transforms U and V are also different. We will use two simple facts.

1. If two complex numbers y and w in (a,b)-format are different, then $|y-w|^2 \geq 2^{-2b}$.
2. If two complex numbers Y and W are such that $|Y-W|^2 \geq 2^{-2b}$, then the $(a,b+1)$-format truncations of Y and W are also different (assuming a large enough to avoid overflow).

We observe now that, by fact (1), $u \neq v$ implies $||u-v||^2 \geq 2^{-2b}$. Thus, $||U-V||^2 = N||u-v||^2 \geq N2^{-2b}$ so that, at least for some $\mathbf{i} \in Z$, $|U(\mathbf{i})-V(\mathbf{i})|^2 \geq 2^{-2b}$, or $|U(\mathbf{i})-V(\mathbf{i})| \geq 2^{-b}$. By fact (2), the latter relation implies that the $(a+\lceil\log_2 N\rceil+1,b+1)$-format truncations of U and V must be different. □

2.2 Note on the VLSI Model of Computation

The results of this work are of interest, in particular, for the *VLSI model of computation*. For a general discussion of this model, the reader is referred to the literature (Abelson & Andreae 1980, Thompson 1980, Brent & Kung 1981, Ullman 1984). Here, we briefly review only some assumptions regarding the input/output (I/O) protocol, as they play a crucial role in the correspondence between area, time, and information exchange.

The I/O protocol is *semellective* (each input is received exactly once), *unilocal* (each input is received at exactly one input port), and *time-* and *place-determinate* (each I/O variable is available in a pre-specified sequence at a pre-specified port, for all instances of the problem). The relation $AT^2 = \Omega\left(I^2\right)$ is valid for unilocal and place-determinate protocols. The relation $A = \Omega\,(I)$ is valid for semellective and time-determinate protocols.

The *word-local* assumption and the *word-serial* assumption are also often considered. An I/O protocol is *word-local* if, for any cut partitioning the chip, $o\,(N)$ input (output) words have some bits entering (exiting) the chip on each side of the cut (Hornick & Sarrafzadeh 1987). An I/O protocol is *word-serial* if, at any time instant, $o\,(N)$ input (output) words have some, but not all, of their bits read (written). Word-local lower bounds to the information exchange in the two-processor setting yield AT^2 lower bounds for word-local VLSI circuits and A lower bounds for word-serial VLSI circuits.

3 Word-Local Lower Bounds for finite Rings and the Complex Field

In this section, we derive a lower bound on the information exchange of the MDDFT and then apply the results mentioned above to obtain AT^2 and A lower bounds for the MDDFT. Occasionally, an AT^2 lower bound has been erroneously claimed in the literature for either the DFT or the MDDFT, as an alleged corollary of Proposition 1 of (Vuillemin 1983), unfortunately stated (Vuillemin 1983, p. 294) for the multiplication of a variable vector by a *fixed* matrix, whereas the proof does assume a *variable* matrix. Indeed, multiplication by a fixed matrix is not generally a transitive function (just consider the identity matrix). In conclusion, Vuillemin's result does not apply to the MDDFT. A lower bound argument must then be specifically tailored to the MDDFT; the argument in this section is based on its invertibility and its time-shift property.

Let P_1 and P_2 be two processors cooperating in the computation of the transform $X(\mathbf{i})$ of the signal $x(\mathbf{j})$, with a word-local input/output protocol. For $h = 1,2$, we define $A_h \triangleq \{\mathbf{j} : x(\mathbf{j})$ *is input by* $P_h\}$ and $B_h \triangleq \{\mathbf{i} : X(\mathbf{i})$ *is output by* $P_h\}$. Our objective is to lower bound I, the number of bits exchanged by P_1 and P_2, under the constraint $|A_1| \geq |A_2| \geq \lfloor|Z|/2\rfloor = \lfloor N/2\rfloor$.

For $A \subseteq Z$, we denote by $r_h(A)$ the number of distinct values taken by the set of transform components $\{X(\mathbf{i}) : \mathbf{i} \in B_h\}$ corresponding to the input signals in the class $C_A \triangleq \{x : x(\mathbf{j}) = 0 \; for \, \mathbf{j} \in Z - A\}$. For $A \subseteq Z$ and $\mathbf{s} \in Z$, we also let $A + \mathbf{s} = \{\mathbf{j}+\mathbf{s} : \mathbf{j} \in A\}$, where the k-th component of the addition $\mathbf{j}+\mathbf{s}$ is taken modulo n_k.

Lemma 1 *For an* $(a,b;a+\lceil\log_2 N\rceil+1,b+1)$-*complex MDDFT, we have that* $r_h(A+\mathbf{s}) \geq 16^{-2N} r_h(A)$, *for any* $A \subseteq Z$ *and* $\mathbf{s} \in Z$.

Proof We say that x_1, x_2 are B_h-*distinguishable* if their truncated transforms satisfy $X'_1(\mathbf{i}) \neq X'_2(\mathbf{i})$, for some $\mathbf{i} \in B_h$. We say that x_1, x_2 are *strongly* B_h-*distinguishable* if their truncated transforms satisfy $|X'_2(\mathbf{i}) - X'_1(\mathbf{i})| \geq 8 \cdot 2^{-b}$, for some $\mathbf{i} \in B_h$. By the definition of $r_h(A)$, C_A contains $r_h(A)$ signals that are pairwise B_h-distinguishable. By a counting argument, for any given signal, there are at most 16^{2N} signals that are *not* strongly B_h-distinguishable from it. Hence, there is a subset $D_A \subseteq C_A$ of at least $16^{-2N} r_h(A)$ signals that are pairwise strongly B_h-

distinguishable. The set

$$D_{A+\mathbf{s}} \triangleq \{y : \textit{there is an } x \in D_A \textit{ such that} \\ y(\mathbf{j}) = x(\mathbf{j} - \mathbf{s}) \textit{ for all } \mathbf{j} \in Z\}$$

is a subset of $C_{A+\mathbf{s}}$. The thesis of the lemma is a consequence of the bound $|D_{A+\mathbf{s}}| \geq 16^{-2N} r_h(A)$ and of the following claim.

Claim The members of $D_{A+\mathbf{s}}$ are pairwise B_h-distinguishable.

Proof of Claim: For $m = 1, 2$, we let x_m be a signal in D_A, X_m its exact transform, and X'_m its truncated transform. Also for $m = 1, 2$, we let y_m be the signal in $D_{A+\mathbf{s}}$ such that $y_m(\mathbf{j}) = x_m(\mathbf{j} - \mathbf{s})$, Y_m the exact transform of y_m, and Y'_m the truncated transform. The goal is to show that the strong B_h-distinguishability of x_1 and x_2 implies the B_h-distinguishability of y_1 and y_2.

The argument relies on the time-shift property of MDDFT: if, for all $\mathbf{j} \in Z$, $y(\mathbf{j}) = x(\mathbf{j} - \mathbf{s})$, then, for all $\mathbf{i} \in Z$, $Y(\mathbf{i}) = X(\mathbf{i})w(\mathbf{i}, \mathbf{s})$, where $w(\mathbf{i}, \mathbf{s}) \triangleq \prod_{k=1}^{d} \omega_k^{-i_k s_k}$. From the expression for $Y(\mathbf{i})$, with some straightforward algebra we obtain

$$\begin{aligned} Y'_2(\mathbf{i}) - Y'_1(\mathbf{i}) = \ & (X'_2(\mathbf{i}) - X'_1(\mathbf{i}))\, w(\mathbf{i}, \mathbf{s}) - \\ & [\, (X'_2(\mathbf{i}) - X_2(\mathbf{i}))\, w(\mathbf{i}, \mathbf{s}) + \\ & (X_1(\mathbf{i}) - X'_1(\mathbf{i}))\, w(\mathbf{i}, \mathbf{s}) + \\ & (Y_2(\mathbf{i}) - Y'_2(\mathbf{i})) + (Y'_1(\mathbf{i}) - Y_1(\mathbf{i}))]. \end{aligned}$$

The latter equation, together with the bounds $|w(\mathbf{i}, \mathbf{s})| = 1$, $|X'_m(\mathbf{i}) - X_m(\mathbf{i})| \leq 2^{-b}$, and $|Y'_m(\mathbf{i}) - Y_m(\mathbf{i})| \leq 2^{-b}$, yields

$$|Y'_2(\mathbf{i}) - Y'_1(\mathbf{i})| \geq |X'_2(\mathbf{i}) - X'_1(\mathbf{i})| - 4 \cdot 2^{-b}.$$

From the strong B_h-distinguishability of x_1 and x_2, there is an $\mathbf{i} \in Z$ such that the right hand side of the above equation is positive, and hence so is the left hand side. In conclusion, y_1 and y_2 are B_h-distinguishable. □

Lemma 2 *Let $A, A_2 \subseteq Z$. There exists an $\mathbf{s} \in Z$ such that $|(A + \mathbf{s}) \cap A_2| \geq |A||A_2|/|Z|$.*

Proof If $\mathbf{j} \in A$, then $\mathbf{j}+\mathbf{s} \in A_2$ for exactly $|A_2|$ values of $\mathbf{s}$. Thus,

$$\sum_{\mathbf{s} \in Z} |(A + \mathbf{s}) \cap A_2| = |A||A_2|.$$

Some $\mathbf{s}$ must achieve an intersection of size no less than the average, which is $|A||A_2|/|Z|$. □

Theorem 1 *For word-local protocols, the information exchange of a $(0, b; \lceil \log_2 N \rceil + 1, b + 1)$-complex MDDFT satisfies the relation $I \geq \frac{b-32}{4} N - \frac{b}{2}$.*

Proof From the invertibility of the transform, it follows that $r_1(A_1) r_2(A_1) \geq 2^{2b|A_1|} \geq 2^{bN}$. If $r_1(A_1) \leq 2^{3bN/4}$, then $r_2(A_1) \geq 2^{bN/4}$. In this case, $I \geq \log_2 r_2(A_1) \geq (b/4)N$, and the theorem is proved.

If $r_1(A_1) > 2^{3bN/4}$, then we argue as follows. From Lemma 2, there is an $\mathbf{s} \in Z$ such that

$$\begin{aligned} |(A_1 + \mathbf{s}) \cap A_2| &\geq |A_1||A_2|/|Z| = \lceil N/2 \rceil \lfloor N/2 \rfloor / N \\ &\geq (N - 1)/4. \end{aligned}$$

Also,

$$\begin{aligned} |(A_1 + \mathbf{s}) \cap A_1| &= |A_1| - |(A_1 + \mathbf{s}) \cap A_2| \\ &\leq \lceil N/2 \rceil - (N - 1)/4 \\ &\leq (N + 1)/4. \end{aligned}$$

From Lemma 1, $r_1(A_1 + \mathbf{s}) \geq 16^{-2N} r_1(A_1) > 2^{(3b-32)N/4}$. We now observe that there are at most $2^{2b \cdot |(A_1+\mathbf{s}) \cap A_1|} \leq 2^{b(N+1)/2}$ ways to choose the values of the $x(\mathbf{j})$'s for $\mathbf{j} \in ((A_1 + \mathbf{s}) \cap A_1)$. Therefore, for at least one such choice, at least $2^{(3b-32)N/4}/2^{b(N+1)/2} = 2^{(b-32)N/4 - b/2}$ distinct output values are produced for the output set $\{X(\mathbf{i}) : \mathbf{i} \in B_1\}$ by changing only the input variables $\{x(\mathbf{j}) : \mathbf{j} \in ((A_1 + \mathbf{s}) \cap A_2)\}$. We can then conclude that $I \geq \frac{b-32}{4} N - \frac{b}{2}$, as desired. □

For simplicity, we have stated Theorem 1 for a $(0, b; \lceil \log_2 N \rceil + 1, b + 1)$-complex MDDFT. By a simple scaling consideration, the result can be easily extended to arbitrary input formats, as long as the output format guarantees invertibility.

Lemma 3 *For the MDDFT on a finite ring R, we have that $r_h(A + \mathbf{s}) = r_h(A)$, for any $A \subseteq Z$ and $\mathbf{s} \in Z$.*

Proof We say that x_1, x_2 are *B_h-distinguishable* if their transforms satisfy $X_1(\mathbf{i}) \neq X_2(\mathbf{i})$ for some $\mathbf{i} \in B_h$. By the definition of $r_h(A)$, there is a set $D_A \subseteq C_A$ of exactly $r_h(A)$ signals that are pairwise B_h-distinguishable. It is easily seen that the set

$$D_{A+\mathbf{s}} \triangleq \{y : \textit{there is an } x \in D_A \textit{ such that} \\ y(\mathbf{j}) = x(\mathbf{j} - \mathbf{s}) \textit{ for all } \mathbf{j} \in Z\}$$

is a subset of $C_{A+\mathbf{s}}$.

Claim The members of $D_{A+\mathbf{s}}$ are pairwise B_h-distinguishable.

This claim is, again, a consequence of the time-shift property: if $y(\mathbf{j}) = x(\mathbf{j} - \mathbf{s})$, then $Y(\mathbf{i}) = X(\mathbf{i})w(\mathbf{i}, \mathbf{s})$, where $w(\mathbf{i}, \mathbf{s}) \triangleq \prod_{k=1}^{d} \omega_k^{-i_k s_k}$. Since all the ω_k's have a multiplicative inverse in R, so does $w(\mathbf{i}, \mathbf{s})$. Hence, if for some $\mathbf{i} \in B_h$ $X_1(\mathbf{i}) \neq X_2(\mathbf{i})$, then $Y_1(\mathbf{i}) \neq Y_2(\mathbf{i})$.

From the above claim, it follows that $r_h(A + \mathbf{s}) \geq r_h(A)$. From a symmetric argument, it follows that $r_h(A) \geq r_h(A + \mathbf{s})$, hence the thesis of the lemma. □

Theorem 2 *For word-local protocols, the information exchange of an $(n_1 \times n_2 \times \ldots \times n_d)$-point MDDFT on a finite ring R satisfies the relation $I \geq \frac{N-2}{8} \log_2 |R|$.*

Proof The proof develops along the same lines of the proof of Theorem 1, with 2^{2b} replaced by $|R|$, and Lemma 3 invoked in place of Lemma 1. □

Since the truncated complex-MDDFT is not linear, in order to maintain the same structure for the proofs of Theorem 1 and Theorem 2, we have not made use of the linearity of the transform even in the case of finite rings. By exploiting linearity along the lines of the argument of Bilardi et al. (1989), the lower bound on I in Theorem 2 can be improved by a constant factor.

In an abstract approach (Nicholson 1971), the Fourier transform for signals $x : G \rightarrow R$ is defined as an isomorphism between $R[G]$, the group algebra of the group G over the ring R, and R^N, the pointwise algebra of N-tuples from R, where N is the order of G. This is an algebraic formulation of the well-known convolution property, and is more general than the definition given in Section 2, Equation 1. In fact, if R decomposes into more than one local ring, then only some of the isomorphisms between $R[G]$ and R^N take the form of Equation 1, as implicit in the work of Dubois & Venetsanopoulos (1978). Theorem 2 can be extended to the more general notion of transform. Omitting the details, we observe that the time-shift property holds in general, with the factor $w(\mathbf{i}, \mathbf{s})$ representing the transform of the signal whose value is 1 when the argument is $\mathbf{s}$, and 0 otherwise.

From Theorems 1 and 2, and from the general relations relating the information exchange to area and time reported in Sections 1 and 2, we obtain the following lower bounds on area and area-time tradeoff.

Theorem 3 *For any VLSI system computing an $(n_1 \times n_2 \times \ldots \times n_d)$-point $(0, b; \lfloor \log_2 N \rfloor + 1, b + 1)$-complex MDDFT, with $b \geq 8$, area and time satisfy the following lower bounds:*

1. *$A = \Omega(Nb)$ if the I/O protocol is word-serial;*
2. *$AT^2 = \Omega(N^2 b^2)$ if the I/O protocol is word-local.*

Theorem 4 *For any VLSI system computing an $(n_1 \times n_2 \times \ldots \times n_d)$-point MDDFT over a finite ring R, area and time satisfy the following lower bounds:*

1. *$A = \Omega(N \log M)$ if the I/O protocol is word-serial;*
2. *$AT^2 = \Omega(N^2 \log^2 M)$ if the I/O protocol is word-local.*

When $R = \mathbb{Z}_M$ and n_1, n_2, …, n_d are relatively prime, the conditions on the existence of the transform mentioned in Section 2 imply that $M \geq N + 1$. Then Theorem 4 implies that $AT^2 = \Omega(N^2 \log^2 N)$. This result was previously established in (Thompson 1979), by different arguments, for the special case $d = 1$, and then generalized to d relatively prime factors in (Hornick & Sarrafzadeh 1988) by an area-time reduction technique (Hornick & Sarrafzadeh 1987) based on the Good-Thomas prime-factor algorithm.

4 Lower Bounds for Finite Fields

In this section, we explore non word-local protocols. We focus on finite fields, which are rings of great interest and with considerable structure. We denote by $\mathrm{GF}(p^m)$ the finite field of p^m elements, where p is a prime integer and m a positive integer, which exists and is unique up to (field) isomorphisms. We consider representations of $\mathrm{GF}(p^m)$ where each element is viewed as a polynomial $a(z) = \sum_{k=0}^{m-1} a_k z^k$ of degree m, specified via the m-tuple of its *coefficients* $(a_0, \ldots, a_{m-1}) \in \mathrm{GF}(p)^m = \mathbb{Z}_p^m$. Polynomial addition $(+)$ and multiplication $(\cdot)$ are taken modulo a given polynomial of degree m, $g(z)$, irreducible over $\mathrm{GF}(p)$. In the following two subsections, we study the information exchange of the DFT over $\mathrm{GF}(p^m)$, for certain classes of non word-local protocols.

4.1 Semi Word-Local Protocols

It is natural to consider, as an intermediate step, a weakening of the word-local constraint by assuming it just for the coefficients rather than for the entire polynomial. Specifically, we say that a protocol is *coefficient-local* if all the bits that represent the same coefficient of any input or output component are handled by the same processor. We begin by presenting a lower bound to the coefficient-local information exchange of the (multidimensional, cyclic) convolution of signals with values in $\mathrm{GF}(p^m)$. We then derive a lower bound for the DFT, under suitably restricted protocols, by making use of both (i) the convolution property of Fourier transforms and (ii) a relation we will establish between the information exchange of an invertible linear transform and that of its inverse. (The inverse Fourier transform intervenes in the convolution property.)

Proposition 2 *Let $u, x : Z \to \mathrm{GF}(p^m)$ two d-dimensional signals. Let P_1 and P_2 be two processors cooperating in the computation of the* cyclic convolution *$v : Z \to \mathrm{GF}(p^m)$ of u and x, defined as*

$$v(\mathbf{i}) = \sum_{\mathbf{j} \in Z} u(\mathbf{j}) \cdot x(\mathbf{i} - \mathbf{j}).$$

Let $x(\mathbf{j}) = \sum_{k=0}^{m-1} x_k(\mathbf{j}) z^k$. Consider a coefficient-local *I/O protocol such that the set of variables*

$$\mathcal{X} = \{x_k(\mathbf{j}) : 0 \leq k < \lfloor m/3 \rfloor \text{ and } \mathbf{j} \in Z\}$$

is equally split between P_1 and P_2. Then, there exist values $h \in \{0, 1, \ldots, 2\lfloor m/3 \rfloor - 1\}$ and $\mathbf{s} \in Z$ such that, if u is fixed as $u(\mathbf{j}) = z^h \delta_{\mathbf{j},\mathbf{s}}$, where δ is the Kronecker symbol, the information exchange satisfies

$$I \geq \frac{1}{2} \left\lfloor \frac{\lfloor m/3 \rfloor N}{2} \right\rfloor \log_2 p. \qquad (2)$$

Proof The proof technique is a variant of shift arguments used for similar problems (Abelson & Andreae 1980, Brent & Kung 1981, Vuillemin 1983). Consider the set of output variables

$$\mathcal{V} = \{v_\ell(\mathbf{i}) : \lfloor m/3 \rfloor \leq \ell < 2\lfloor m/3 \rfloor \text{ and } \mathbf{i} \in Z\}.$$

It is easily seen that if $x_k(\mathbf{j}) \in \mathcal{X}$ and $v_\ell(\mathbf{i}) \in \mathcal{V}$, then setting $u(\mathbf{j}) = z^{\ell-k} \delta_{\mathbf{j},\mathbf{i}}$ yields $v_\ell(\mathbf{i}) = x_k(\mathbf{j})$. Since $0 \leq \ell - k < 2\lfloor m/3 \rfloor$, we have a set S of $2\lfloor m/3 \rfloor N$ signals among which to select u. Let now $\mathcal{X}_1$ and $\mathcal{X}_2$ the subsets of $\mathcal{X}$ respectively handled by P_1 and P_2. Similarly define $\mathcal{V}_1$ and $\mathcal{V}_2$. As u varies in S, each pair of coefficients in $(\mathcal{X}_1 \times \mathcal{V}_2) \cup (\mathcal{X}_2 \times \mathcal{V}_1)$ contributes exactly once to the information exchange I_u of the convolution with fixed u and variable x. Hence, considering that the information content of one coefficient is $\log_2 p$ bits, we have:

$$\sum_{u \in S} I_u \geq |(\mathcal{X}_1 \times \mathcal{V}_2) \cup (\mathcal{X}_2 \times \mathcal{V}_1)| \log_2 p$$
$$\geq (N \lfloor \lfloor m/3 \rfloor / 2 \rfloor)(N \lfloor m/3 \rfloor) \log_2 p,$$

where we have considered that $|\mathcal{X}_1|, |\mathcal{X}_2| \geq N\lfloor \lfloor m/3 \rfloor / 2 \rfloor$ and that $|\mathcal{V}_1| + |\mathcal{V}_2| = N\lfloor m/3 \rfloor$. Dividing by the cardinality of S we conclude that, on average, the information exchange is at least $\frac{1}{2} \left\lfloor \frac{\lfloor m/3 \rfloor N}{2} \right\rfloor \log_2 p$, whence the statement of the proposition. □

A first, interesting corollary of the preceding proposition is an area-time lower bound for the convolution itself.

Theorem 5 *For any VLSI system computing an $(n_1 \times n_2 \times \ldots \times n_d)$-point convolution over $\mathrm{GF}(p^m)$, with a coefficient-local protocol, $AT^2 = \Omega((Nm \log p)^2)$.*

Next, we show that, in vector spaces over a finite field, a linear transform and its inverse have nearly the same information exchange, for suitable I/O protocols. This result will be applied to the DFT over $\mathrm{GF}(p^m)$, regarded as a linear transform (on the vector space $\mathrm{GF}(p)^{Nm}$) between the input coefficients $\{x_k(\mathbf{j}) : 0 \le k < m \text{ and } \mathbf{j} \in Z\}$ and the output coefficients $\{X_\ell(\mathbf{i}) : 0 \le \ell < m \text{ and } \mathbf{i} \in Z\}$.

Proposition 3 *Let* $\mathbf{F}$ *be a finite field and let* $\mathcal{T} : \mathbf{F}^q \to \mathbf{F}^q$ *be an* invertible linear transform*. Let the input (resp., output) vector be partitioned into two subvectors* w_1 *and* w_2 *(resp.,* W_1 *and* W_2*). Consider an I/O protocol where* w_1 *and* W_1 *are handled by* P_1 *and* w_2 *and* W_2 *are handled by* P_2*. Then, the information exchange* $I_{\mathcal{T}^{-1}}$ *of the inverse transform is related to that of the direct transform* $I_{\mathcal{T}}$ *as*

$$I_{\mathcal{T}^{-1}} \le 2 I_{\mathcal{T}}.$$

Proof We begin by writing $\mathcal{T}$ in block-matrix form:

$$W_1 = T_{11} w_1 + T_{12} w_2$$
$$W_2 = T_{21} w_1 + T_{22} w_2$$

It is easy to argue (see also (Thompson 1979)) that, to compute the above equations, it is necessary and sufficient that P_1 receive from P_2 information specifying $T_{21} w_1$ and that P_2 receive from P_1 information specifying $T_{12} w_2$. Let $r(A)$ denote the rank of matrix A. Since there are exactly $|\mathbf{F}|^{r(A)}$ distinct vectors in the image of A, we have:

$$I_{\mathcal{T}} = (r(T_{12}) + r(T_{21})) \log_2 |\mathbf{F}|. \qquad (3)$$

To bound $I_{\mathcal{T}^{-1}}$ from above, we assume now that P_1 is given W_1, P_2 is given W_2, and develop an algorithm such that P_1 produces w_1 and P_2 produces w_2. We make use of the following straightforward definitions and relations:

$$(w_1^1, w_2^1) \triangleq \mathcal{T}^{-1}(W_1, 0)$$
$$(w_1^2, w_2^2) \triangleq \mathcal{T}^{-1}(0, W_2)$$
$$w_1 = w_1^1 + w_1^2$$
$$w_2 = w_2^1 + w_2^2$$

Clearly, P_1 can autonomously compute w_1^1, but must receive w_1^2 from P_2. A compact encoding of w_1^2 can be obtained by observing that, as a consequence of the definition of w_1^2, we have $T_{11} w_1^2 + T_{12} w_2^2 = 0$ or, equivalently,

$$T_{11} w_1^2 = -T_{12} w_2^2.$$

The right-hand side of the above equation belongs to the range of T_{12}, which can be encoded via $r(T_{12})$ elements of $\mathbf{F}$, by providing its representation in a chosen basis of such range. This identifies w_1^2 up to an additive vector belonging to the kernel of T_{11}, hence it can be encoded via $\nu(T_{11})$ elements of $\mathbf{F}$, where $\nu(A)$ denotes the dimension of the null space of matrix A. In summary, it is sufficient for P_2 to send P_1 a number $r(T_{12}) + \nu(T_{11})$ of $\mathbf{F}$ elements. A symmetric argument applies to the information flowing in the other direction, leading to the bound

$$I_{\mathcal{T}^{-1}} \le [r(T_{12}) + \nu(T_{11}) + r(T_{21}) + \nu(T_{22})] \log_2 |\mathbf{F}|. \qquad (4)$$

Since $\mathcal{T}$ is invertible, any subset of columns of its matrix is linearly independent. Then, from standard linear algebra results, we obtain the following relations, where q_1 (q_2) is the dimension of w_1 (w_2):

$$\begin{aligned} r(T_{11}) + \nu(T_{11}) &= q_1 \\ r(T_{22}) + \nu(T_{22}) &= q_2 \\ r(T_{11}) + r(T_{21}) &\ge q_1 \\ r(T_{22}) + r(T_{12}) &\ge q_2 \end{aligned}$$

By subtracting the sum of the first two equations from the sum of the latter two ones, we obtain the relation

$$r(T_{21}) + r(T_{12}) \ge \nu(T_{11}) + \nu(T_{22}),$$

which, in combination with Equation 4, yields

$$I_{\mathcal{T}^{-1}} \le 2(r(T_{12}) + r(T_{21})) \log_2 |\mathbf{F}|,$$

which, together with Equation 3, yields the bound stated by this proposition. □

We are now ready to tackle the DFT. The approach consists in realizing a convolver based on the computation of the DFT and its inverse, thus linking the information exchange of the DFT to that of the convolution, for which we have derived a lower bound earlier in this section. For reasons that will become apparent in the context of the proof, we consider coefficient-local protocols where all the output coefficients of the same word are handled by the same processor, and call them *semi word-local.* In other words, in a semi word-local protocol the output is word-local but the input is just coefficient-local.

Theorem 6 *For* semi word-local protocols*, the information exchange of an* $(n_1 \times n_2 \times \ldots \times n_d)$*-point MDDFT on a finite field* $\mathrm{GF}(p^m)$ *satisfies the relation* $I = \Omega(N \log |\mathrm{GF}(p^m)|) = \Omega(Nm \log p)$*. Correspondingly,* $AT^2 = \Omega\left((Nm \log p)^2\right)$*.*

Proof We denote by x the input of the DFT and let $x(\mathbf{j}) = \sum_{k=0}^{m-1} x_k(\mathbf{j}) z^k$. As in the proof of Proposition 2, we consider a *coefficient-local* I/O protocol such that the set of variables

$$\mathcal{X} = \{x_k(\mathbf{j}) : 0 \le k < \lfloor m/3 \rfloor \text{ and } \mathbf{j} \in Z\}$$

is equally split between P_1 and P_2. If v is the convolution of u and x, then by the convolution property, the transforms satisfy

$$V(\mathbf{j}) = U(\mathbf{j}) \cdot X(\mathbf{j}) \qquad \forall \mathbf{j} \in Z. \qquad (5)$$

We now fix u to be the signal whose existence is established in Proposition 2. Processors P_1 and P_2 can then compute v by the following steps. Observe that, since u is fixed, its transform U can be precomputed off-line and "hardwired" into the algorithm.

1. Compute the transform X of x.
2. Compute the transform V by Equation 5.
3. Compute the convolution v by antitrasforming V.

The information exchange is then bounded by I_{DFT} in the first step; it is 0 in the second step, since, for each $\mathbf{j}$, both $U(\mathbf{j})$ are $X(\mathbf{j})$ are handled by the same processor; it is bounded by $I_{DFT^{-1}} \le 2 I_{DFT}$ in the third step (applying Proposition 3 to the DFT, with $\mathbf{F} = \mathrm{GF}(p)$ and $q = Nm$). Thus, overall, the information exchange I to compute the convolution by fixed u is bounded as

$$I \le 3 I_{DFT}.$$

Combining the latter with Equation 2, we get

$$I_{DFT} \ge \frac{1}{6} \left\lfloor \frac{\lfloor m/3 \rfloor N}{2} \right\rfloor \log_2 p,$$

whence the stated result. □

4.2 A non Word-Local Protocol

To gain some understanding on general protocols, we need to investigate how information moves across different sets of coefficients of the field elements. In this section, we begin with the simplest case where $R = \mathrm{GF}\left(p^2\right)$, with p an odd prime. Then, the input $x : Z \to \mathrm{GF}\left(p^2\right)$ and the output $X : Z \to \mathrm{GF}\left(p^2\right)$ of the transform are vectors of linear polynomials whose components can be written as

$$\begin{aligned} x(\mathbf{j}) &= x_0(\mathbf{j}) + x_1(\mathbf{j}) \cdot z, \\ X(\mathbf{i}) &= X_0(\mathbf{i}) + X_1(\mathbf{i}) \cdot z. \end{aligned}$$

We focus on the two-processor protocol where, for each $\mathbf{i}, \mathbf{j} \in Z$, P_1 handles the coefficients $x_0(\mathbf{j})$ and $X_0(\mathbf{i})$, and P_2 handles $x_1(\mathbf{j})$ and $X_1(\mathbf{i})$. We will establish that $I = \Omega\left(N \log p\right)$ by showing how to view each element of $\mathrm{GF}\left(p^2\right)$ as the sum of a "real" and an "imaginary" part, pretty much as with complex numbers. Furthermore, we will introduce an involutory permutation σ over the signal domain Z, and the related concepts of even (odd) signal, which is left unchanged (changes sign) under σ. As it turns out, the transform of an odd, real signal is odd and imaginary, thus transferring information from the real to the imaginary part, which are closely related (though not exactly equal) to the first and the second coefficient of the field elements. For the formal developments, we need a number of concepts introduced below.

Definition 1 *The* conjugate *of element $a \in \mathrm{GF}\left(p^2\right)$ is $a^* \triangleq a^p$. We say that $a \in \mathrm{GF}\left(p^2\right)$ is* real *if $a^* = a$, and is* imaginary *if $a^* = -a$.*

It can be shown that a and b are conjugate according to Definition 1 if and only if they are conjugate according to the standard field-theoretic definition, *i.e.*, a and b have the same minimal polynomial.

Lemma 4 *Let $a, b \in \mathrm{GF}\left(p^2\right)$. Then, we have:*

- A: $(a^*)^* = a$.
- B: $(a + b)^* = a^* + b^*$.
- C: $(a \cdot b)^* = a^* \cdot b^*$.
- D: *If a and b are both real, then $a + b$ and $a \cdot b$ are real.*
- E: *If a and b are both imaginary, then $a + b$ is imaginary and $a \cdot b$ is real.*
- F: *If a is real and b is imaginary, then $a \cdot b$ is imaginary.*
- G: *a is both real and imaginary if and only if $a = 0$.*

Proof A: we have $(a^*)^* = a^{p^2}$ and $a^{p^2-1} = 1$, hence the thesis. B: in any finite field of characteristic p, $(a+b)^p = a^p + b^p$. C: straightforward from $(a \cdot b)^p = a^p \cdot b^p$.
D,E,F (addition/multiplication): from B/C and the definition of real and imaginary elements.
G: If $a = 0$, then $a^* = 0^p = 0 = a = -a$, hence a is both real and imaginary. Conversely, if $a^* = a = -a$, then $a = 0$ since $p > 2$. □

Lemma 5 *Every element $a \in \mathrm{GF}\left(p^2\right)$ can be uniquely expressed as the sum of a real element and an imaginary element.*

Proof Clearly, $a = \Re(a) + \Im(a)$, where

$$\Re(a) \triangleq (a + a^*)/2, \qquad \Im(a) \triangleq (a - a^*)/2.$$

By applying Properties A and B of Lemma 4 it can be shown that $\Re(a)$ is real and $\Im(a)$ is imaginary. To prove that the representation is unique, suppose that $a = a_R + a_I = a_{R'} + a_{I'}$, where a_R and $a_{R'}$ are real and a_I and $a_{I'}$ imaginary. We can then write

$$a_{R'} + (-1) \cdot a_R = a_I + (-1) \cdot a_{I'}.$$

Now, (-1) is real since $(-1)^* = (-1)^p = -1$. In the above equality, the right hand side is real (by D) and the left hand side is imaginary (by F and E). Then, by G, $a_{R'} - a_R = a_I - a_{I'} = 0$, hence the two representations for a must be equal. □

The next lemma provides us with valuable properties of real and imaginary elements, in terms of any polynomial representation of $\mathrm{GF}\left(p^2\right)$.

Lemma 6 *In $\mathrm{GF}\left(p^2\right)$, there are exactly p real elements and p imaginary elements. An element $a = a_0 + a_1 \cdot z$ of $\mathrm{GF}\left(p^2\right)$ is real if and only if $a_1 = 0$. Two imaginary elements $b = b_0 + b_1 \cdot z$ and $c = c_0 + c_1 \cdot z$ are equal if and only if $b_1 = c_1$.*

Proof Since the real elements are, by definition, roots of a polynomial of degree p, they are at most p in number; the same argument holds for the imaginary elements. Then, there must be exactly p of each, since their pairs do generate all p^2 elements of $\mathrm{GF}\left(p^2\right)$, by Lemma 5.

The real elements coincide with the subfield of field integers $\mathrm{GF}\left(p\right)$. In fact, the multiplicative identity is real ($1^p = p$) and so must be its p integer multiples, by Property D of Lemma 4. The identity is represented by $1 + 0 \cdot z$, hence the integer obtained by summing a_0 identity terms is represented by $a_0 + 0 \cdot z$.

Let b and c be imaginary. Obviously, if $b = c$, then $b_1 = c_1$. Conversely, if $b_1 = c_1$, then $c - b = (c_0 - b_0) + 0 \cdot z$, whence $c - b = 0$, since the result is simultaneously real (no linear coefficient) and imaginary (the difference of two imaginary elements). □

Next, we introduce some classes of signals with respect to which the Fourier transform exhibits properties useful to investigate the information exchange.

Definition 2 *Let σ be the permutation of the signal domain such that, for every $\mathbf{j} = (j_1, \ldots, j_d) \in Z$,*

$$\sigma(\mathbf{j}) \triangleq (p \cdot j_1 \mod n_1, \ldots, p \cdot j_d \mod n_d).$$

A signal $x : Z \to \mathrm{GF}\left(p^2\right)$ is even *if $x(\sigma(\mathbf{j})) = x(\mathbf{j})$, and is* odd *if $x(\sigma(\mathbf{j})) = -x(\mathbf{j})$, $\forall \mathbf{j} \in Z$.*

Now, consider the entries of the Fourier matrix $w(\mathbf{i}, \mathbf{j}) \triangleq \omega_1^{-i_1 j_1} \omega_2^{-i_2 j_2} \cdots \omega_d^{-i_d j_d}$, where, as usual, $\mathbf{i} = (i_1, \ldots, i_d) \in Z$ and $\mathbf{j} = (j_1, \ldots, j_d) \in Z$. A key property relates σ to $w(\mathbf{i}, \mathbf{j})$, *i.e.*:

$$w(\mathbf{i}, \sigma(\mathbf{j})) = w^*(\mathbf{i}, \mathbf{j}).$$

To see that the property holds, just recall Definition 2 for σ and observe that

$$\omega_k^{-i_k (p j_k \mod n_k)} = \omega_k^{-i_k (p j_k)} = (\omega_k^{-i_k j_k})^p = (\omega_k^{-i_k j_k})^*.$$

We say that x is real (resp., imaginary) if all its components are real (imaginary). We introduce four classes of signals: *real and even* (RE), *real and odd* (RO), *imaginary and even* (IE), *imaginary and odd* (IO). The MDDFT of these classes enjoys properties similar to those well known in the complex field.

Lemma 7 *Let X be the MDDFT of $x : Z \rightarrow \mathrm{GF}\left(p^2\right)$.*

- A: *if x is RE, then X is RE.*
- B: *if x is RO, then X is IO.*
- C: *if x is IE, then X is IE.*
- D: *if x is IO, then X is RO.*

Proof By summing expression 1 for the MDDFT with its version resulting from permuting the terms of the summation by σ, we derive the expression

$$\begin{aligned} X(\mathbf{i}) &= \frac{1}{2}\sum_{\mathbf{j}\in Z}[x(\mathbf{j})w(\mathbf{i},\mathbf{j}) + x(\sigma(\mathbf{j}))w(\mathbf{i},\sigma(\mathbf{j}))] \\ &= \frac{1}{2}\sum_{\mathbf{j}\in Z}[x(\mathbf{j})w(\mathbf{i},\mathbf{j}) + x(\sigma(\mathbf{j}))w^*(\mathbf{i},\mathbf{j})]. \end{aligned}$$

In a similar fashion, one obtains

$$X(\sigma(\mathbf{i})) = \frac{1}{2}\sum_{\mathbf{j}\in Z}[x(\mathbf{j})w^*(\mathbf{i},\mathbf{j}) + x(\sigma(\mathbf{j}))w(\mathbf{i},\mathbf{j})].$$

When x is even, X is clearly even as well and it can be rewritten as

$$X(\mathbf{i}) = \sum_{\mathbf{j}\in Z} x(\mathbf{j})\Re\left(w(\mathbf{i},\mathbf{j})\right).$$

If x is also real, then so is X by Property D of Lemma 4. Similarly, if x is imaginary then X is imaginary as well. This concludes the proof of A and C. When x is odd, X is odd and can be rewritten as

$$X(\mathbf{i}) = \sum_{\mathbf{j}\in Z} x(\mathbf{j})\Im\left(w(\mathbf{i},\mathbf{j})\right).$$

If x is also real, then X is imaginary; if x is odd, then X is real (proving B and D). □

We are now ready to combine concepts and observations introduced in this section to arrive at its main result.

Theorem 7 *Let $X(\mathbf{i}) = X_0(\mathbf{i}) + X_1(\mathbf{i}) \cdot z$ be the $\mathbf{i}$-th component of the $(n_1 \times n_2 \times \ldots \times n_d)$-point MDDFT of the signal $x : Z \rightarrow \mathrm{GF}\left(p^2\right)$ with $\mathbf{j}$-th component $x(\mathbf{j}) = x_0(\mathbf{j}) + x_1(\mathbf{j}) \cdot z$.*

Let I be the information exchange of a two-processor system where, for each $\mathbf{i}, \mathbf{j} \in Z$, P_1 handles $x_0(\mathbf{j})$ and $X_0(\mathbf{i})$, and P_2 handles $x_1(\mathbf{j})$ and $X_1(\mathbf{i})$.

If each n_k divides $p-1$, then $I = 0$. Otherwise, $I \geq \frac{N}{4}\log_2 p$.

Proof If n_k divides $p-1$ for $1 \leq k \leq d$, consider a primitive $(p-1)$-st root of unity α and observe that α^{h_k}, with $h_k \triangleq (p-1)/n_k$, is a primitive n_k-th root of unity: in fact, it is straightforward to see that α^0, α^{h_k}, α^{2h_k}, ..., $\alpha^{(n_k-1)h_k}$ are distinct (the exponents are all less than $p-1$) roots of unity, and $\alpha^{n_k h_k} = \alpha^{p-1} = 1$. Since α is real (see the considerations on real elements in the proof of Lemma 6), all the n_k-th roots of unity are real and belong to the subfield $\mathrm{GF}(p)$ of field integers. The property is true for $1 \leq k \leq d$, hence $w(\mathbf{i},\mathbf{j}) \in \mathrm{GF}(p)$ $\forall \mathbf{i},\mathbf{j} \in Z$ and Equation 1 for the MDDFT can be rewritten as

$$X_0(\mathbf{i}) + X_1(\mathbf{i}) \cdot z = \sum_{\mathbf{j}\in Z} x_0(\mathbf{j})w(\mathbf{i},\mathbf{j}) + z \cdot \sum_{\mathbf{j}\in Z} x_1(\mathbf{j})w(\mathbf{i},\mathbf{j}).$$

In other words, $X_0(\mathbf{i})$ can be computed when only the x_0 coefficients are known, and $X_1(\mathbf{i})$ can be independently computed with the sole knowledge of the x_1 coefficients. Since P_1 handles $x_0(\mathbf{j})$ and $X_0(\mathbf{i})$, and P_2 handles $x_1(\mathbf{j})$ and $X_1(\mathbf{i})$, we conclude that P_1 and P_2 can produce their output without exchanging any information, whence $I = 0$.

If there exists $\bar{k}$ such that $n_{\bar{k}}$ does not divide $p-1$, we observe that, when x varies among the real and odd signals, P_2 receives no information on x from its inputs, as $x_1(\mathbf{j}) = 0$ for all $\mathbf{j}$'s. On the other hand, by Property B of Lemma 7, X is imaginary and, by Lemma 6, it is fully reconstructable from its X_1 coefficients, hence X (and consequently its inverse transform x) must essentially be known to P_2 by the end of the computation. Then, the information transfer from P_1 to P_2 satisfies the bound

$$I \geq \log_2(p^{(N-\zeta(n_1,n_2,\ldots,n_d))/2}), \tag{6}$$

where the argument of the logarithm is the number of distinct signals in RO, expressed in terms of the number $\zeta(n_1, n_2, \ldots, n_d)$ of input components that must be null in any odd signal. We are then left with the task of bounding ζ from above. Clearly, ζ gives the number of fixed points of permutation σ. Since a fixed point must leave unchanged all d coordinates of a point in Z, our function can be factored as

$$\zeta(n_1, n_2, \ldots, n_d) = \zeta(n_1)\zeta(n_2)\ldots\zeta(n_d),$$

where $\zeta(n)$, $1 \leq n \leq p^2 - 1$, can be viewed as the number of n-th roots of unity that are real: in fact, $a \in \mathrm{GF}\left(p^2\right)$ is real if and only if $a^* = a$.

Clearly, $\zeta(n) \leq n$, for any n. Furthermore, we will prove below that, if n does not divide $p-1$, then $\zeta(n) \leq n/2$. Using the latter bound on $\zeta(n_{\bar{k}})$ and the former bound on all other factors, the preceding equation yields

$$\zeta(n_1, n_2, \ldots, n_d) \leq n_1 n_2 \cdots n_d/2 = N/2.$$

When plugged into Equation 6, the latter bound results in $I \geq \frac{N}{4}\log_2 p$, as in the theorem statement.

It remains to prove that, if n does not divide $p-1$, then $\zeta(n) \leq n/2$. A real n-th root of unity must also be a $(p-1)$-st root, whence it must also be a g-th root, with $g \triangleq \mathrm{GCD}(n, p-1)$. Then $n = fg$, with $f \geq 2$, because n does not divide $p-1$. In conclusion, $\zeta(n) \leq g = n/f \leq n/2$, as claimed. □

A variant of Theorem 7 where the output coefficients are swapped between the two processors can be similarly proven, based on RE input signals.

5 Conclusions

In this paper we have extended the understanding of the information exchange and the related area-time tradeoffs for the computation of the DFT. The main contributions lie in two direction. First, we have developed bounds for the complex DFT, which escaped previous analysis. Second, we have made some progress towards removing the word-local assumption in the I/O protocol. The lower bounds of Sections 3, 4.1, and 4.2 all exploit different properties of the DFT which are, respectively, the time-shift, the convolution, and the conjugate symmetry properties, thus shading light from different perspectives.

Considerable work remains to be done before the subject can be considered well understood. The full removal of the word-local assumption appears still challenging and is likely to require the exploitation of deeper algebraic properties of the ring where the signals and their transform take values.

References

Abelson, H. & Andreae, P. (1980), ‘Information transfer and area-time tradeoffs for VLSI multiplication’, *Communications of the ACM* **23**(1), 20.

Agarwal, R. C. & Burrus, C. S. (1975), ‘Number theoretic transforms to implement fast digital convolution’, *Proceedings of the IEEE* **63**(4), 550–560.

Alnuweiri, H. M. (1994), ‘Optimal VLSI networks for multidimensional transforms’, *IEEE Transactions on Parallel and Distributed Systems* **5**(7), 763–769.

Bilardi, G., Hornick, S. W. & Sarrafzadeh, M. (1989), Optimal VLSI architectures for multidimensional DFT, *in* ‘Annual ACM Symposium on Parallel Algorithms and Architectures’, ACM, Santa Fe, New Mexico, United States, pp. 265–272.

Bilardi, G. & Preparata, F. P. (1985), The influence of key length on the area-time complexity of sorting, *in* ‘Automata, Languages and Programming’, Springer-Verlag, pp. 53–62.

Bilardi, G. & Preparata, F. P. (1986), ‘Area-time lower-bound techniques with applications to sorting’, *Algorithmica* **1**(1-4), 65–91.

Bilardi, G. & Sarrafzadeh, M. (1987), ‘Optimal VLSI Circuits for Discrete Fourier Transform’, *Advances in Computing Research* **4**, 87–101.

Blahut, R. E. (1983), *Theory and Practice of Error Control Codes*, Addison-Wesley.

Bonneau, R. J. (1973), ‘A class of finite computation structures supporting the fast Fourier transform’.

Brent, R. P. & Kung, H. T. (1981), ‘The Area-Time Complexity of Binary Multiplication’, *Journal of the ACM* **28**(3), 521.

Chakrabarti, C. & JáJá, J. (1991), ‘VLSI architectures for multidimensional transforms’, *IEEE Transactions on Computers* **40**(9), 1053–1057.

Chowdary, N. & Steenaart, W. (1984), ‘A high speed two-dimensional FFT processor’, *IEEE International Conference on Acoustics, Speech, and Signal Processing* pp. 177–180.

Cole, R. & Siegel, A. (1988), ‘Optimal VLSI circuits for sorting’, *Journal of the ACM* **35**(4), 777.

Dubois, E. & Venetsanopoulos, A. N. (1978), ‘The Discrete Fourier Transform Over Finite Rings with Application to Fast Convolution’, *IEEE Transactions on Computers* **C-27**(7), 586–593.

Dubois, E. & Venetsanopoulos, A. N. (1980), ‘The generalized discrete Fourier transform in rings of algebraic integers’, *IEEE Transactions on Acoustics, Speech and Signal Processing* **28**(2), 169–175.

Duris, P., Sykora, O., Vrt’o, I. & Thompson, C. D. (1985), ‘Tight chip area lower bounds for discrete Fourier and Walsh-Hadamard transformations’, *Information Processing Letters* **21**(5), 245–247.

Fedorenko, S. V. (2006), ‘A method for computation of the discrete Fourier transform over a finite field’, *Problems of Information Transmission* **42**(2), 139–151.

Gertner, I. & Shamash, M. (1987), ‘VLSI Architectures for Multidimensional Fourier Transform Processing’, *IEEE Transactions on Computers* **C-36**(11), 1265–1274.

Hornick, S. W. & Sarrafzadeh, M. (1987), ‘On problem transformability in VLSI’, *Algorithmica* **2**(1-4), 97–111.

Hornick, S. W. & Sarrafzadeh, M. (1988), ‘Optimal multidimensional DFT in VLSI’.

Lidl, R. & Pilz, G. (1998), *Applied abstract algebra*, second edn, Springer.

Marino, F. & Swartzlander, E. E. (1999), ‘Parallel implementation of multidimensional transforms without interprocessor communication’, *IEEE Transactions on Computers* **48**(9), 951–960.

Nicholson, P. J. (1971), ‘Algebraic theory of finite Fourier transforms’, *Journal of Computer and System Sciences* **5**(5), 524–547.

Pollard, J. M. (1971), ‘The fast Fourier transform in a finite field’, *Mathematics of Computation* **25**(114), 365–365.

Preparata, F. P. & Sarwate, D. V. (1977), ‘Computational complexity of Fourier transforms over finite fields’, *Mathematics of Computation* **31**(139), 740–740.

Thompson, C. D. (1979), Area-time complexity for VLSI, *in* ‘Annual ACM Symposium on Theory of Computing’, ACM, Atlanta, Georgia, United States, pp. 81–88.

Thompson, C. D. (1980), A Complexity Theory for VLSI, Thesis (phd), Carnegie-Mellon University.

Thompson, C. D. (1983), ‘Fourier Transforms in VLSI’, *IEEE Transactions on Computers* **C-32**(11), 1047–1057.

Ullman, J. D. (1984), *Computational Aspects of VLSI*, Computer Science Press.

Vuillemin, J. (1983), ‘A Combinatorial Limit to the Computing Power of VLSI Circuits’, *IEEE Transactions on Computers* **C-32**(3), 294–300.

Wang, Y. & Zhu, X. (1988), ‘A fast algorithm for the Fourier transform over finite fields and its VLSI implementation’, *IEEE Journal on Selected Areas in Communications* **6**(3), 572–577.

Wu, I.-C. (1984), Area-time tradeoffs in VLSI algorithms, Thesis (msc), National Taiwan University.

Yeh, C.-H. (2002), $AT^2L^2 \approx N^2/2$ for fast Fourier transform in multilayer VLSI, *in* ‘Annual ACM Symposium on Parallel Algorithms and Architectures’, ACM Press, New York, New York, USA, p. 145.