A New Algorithm for Fast Generalized DFTs

CHLOE CHING-YUN HSU and CHRIS UMANS, Caltech, USA

We give an new arithmetic algorithm to compute the generalized Discrete Fourier Transform (DFT) over finite groups G. The new algorithm uses $O(|G|^{\omega/2+o(1)})$ operations to compute the generalized DFT over finite groups of Lie type, including the linear, orthogonal, and symplectic families and their variants, as well as all finite simple groups of Lie type. Here ω is the exponent of matrix multiplication, so the exponent $\omega/2$ is optimal if $\omega = 2$.

Previously, "exponent one" algorithms were known for supersolvable groups and the symmetric and alternating groups. No exponent one algorithms were known, even under the assumption $\omega = 2$, for families of linear groups of fixed dimension, and indeed the previous best-known algorithm for $\mathrm{SL}_2(\mathbb{F}_q)$ had exponent 4/3 despite being the focus of significant effort. We unconditionally achieve exponent at most 1.19 for this group and exponent one if $\omega = 2$.

Our algorithm also yields an improved exponent for computing the generalized DFT over general finite groups G, which beats the longstanding previous best upper bound for any ω. In particular, assuming $\omega = 2$, we achieve exponent $\sqrt{2}$, while the previous best was 3/2.

CCS Concepts: • **Theory of computation → Theory and algorithms for application domains**;

Additional Key Words and Phrases: Discrete Fourier transform, finite groups of Lie type

ACM Reference format:
Chloe Ching-Yun Hsu and Chris Umans. 2019. A New Algorithm for Fast Generalized DFTs. *ACM Trans. Algorithms* 16, 1, Article 4 (November 2019), 20 pages.
https://doi.org/10.1145/3301313

1 INTRODUCTION

Let G be a finite group, and let $\mathrm{Irr}(G)$ denote a complete set of irreducible representations. Given an element c of the group algebra $\mathbb{C}[G]$, a generalized DFT is a linear transform that takes c to

$$\sum_{g \in G} c_g \cdot \bigoplus_{\rho \in \mathrm{Irr}(G)} \rho(g).$$

This is the fundamental linear operation that maps the standard basis for the group algebra $\mathbb{C}[G]$ to a "Fourier basis" of irreducible representations of the group G (which is specified in advance). It has applications in data analysis [19], as a component in other algorithms (including fast operations

A version of this article appeared in SODA 2018 as Reference [9].
We thank the SODA 2018 and Transactions on Algorithms referees for their careful reading of this article and many useful comments. This work is supported by the National Science Foundation under grant number CCF-1423544 and a Simons Foundation Investigator Award.
Authors' address: C. C.-Y. Hsu and C. Umans, Caltech, 1200 E. California Blvd. Pasadena, CA, 91125; emails: chloehsu@berkeley.edu, umans@cs.caltech.edu.

1549-6325/2019/11-ART4 $15.00
https://doi.org/10.1145/3301313

on polynomials and in the Cohn-Umans matrix multiplication algorithms), and as the basis for quantum algorithms for problems entailing a Hidden Subgroup Problem [17]. As one varies the underlying group G, the generalized DFT is a rich source of structured linear maps, which one can hope to compute in nearly-linear time, generalizing the famous Cooley-Tukey FFT for cyclic groups of order 2^k.

We typically speak of the complexity of computing this map in the (non-uniform) arithmetic circuit model and do not concern ourselves with *finding* the irreducible representations. The trivial algorithm thus requires $O(|G|^2)$ operations. The best-known algorithm that works for general finite groups G achieves $O(|G|^{1.5})$ operations,[1] assuming the exponent of matrix multiplication is 2 (see Section 2). For a number of special cases, "exponent one" algorithms are known: For Abelian groups, the symmetric and alternating groups [5, 16], and the so-called *supersolvable* groups [1]. A group that has resisted such exponent one algorithms despite a significant amount of work is $\mathrm{SL}_2(\mathbb{F}_q)$, where the best-known algorithm achieves $O(|G|^{4/3})$ [10]. This group was described as a "particularly interesting and thorny special case" by Maslen, Rockmore, and Wolff [14].

In this article, we obtain exponent one for $\mathrm{SL}_2(\mathbb{F}_q)$ under the assumption that $\omega = 2$ (ω is the exponent of matrix multiplication). Using the current best upper bound $\omega < 2.3729$ [11], we obtain exponent 1.19 for $\mathrm{SL}_2(\mathbb{F}_q)$ unconditionally, which improves the previous 4/3 exponent. Our new algorithm is quite general and leads to a broad array of new results:

- We achieve exponent $\omega/2$ for essentially all linear groups including the general, orthogonal, and symplectic groups, and their special and projective versions, and for all finite groups with a *split* (B, N)*-pair*; we work out the most common cases explicitly in this article in Section 5.
- We achieve exponent $\omega/2$ for all finite simple groups (see Theorem 5.8).
- We achieve an exponent bound for general groups G, which beats the longstanding previous best upper bound, when using exponent-α matrix multiplication as a black box, for any α (see Theorem 6.2). To do this, we prove a structural result about arbitrary finite groups (Theorem 6.1) that relies on the Classification Theorem, which may be of independent interest. In particular, assuming $\omega = 2$, we achieve exponent $\sqrt{2}$, while the previous best was 3/2.

The main idea. At its core, the seminal Beth-Clausen fast generalized DFT is a recursive algorithm that computes a DFT with respect to G by computing several DFTs with respect to H, a subgroup of G. Each of the $[G : H]$ many H-DFTs is lifted to G and then summed together. See Corollary 2.2. A bottleneck in this algorithm comes from the final summation step, which in general costs $[G : H]|G|$. Since there are groups whose largest subgroup H has index at least $|G|^{1/2}$, exponent 3/2 is the best general result possible within this approach. Improvements have generally come from using specific knowledge of how the induced representations from H up to G break up; this can sometimes be used to circumvent the bottleneck summation. In the case of supersolvable groups and the symmetric and alternating groups, this has yielded exponent one algorithms [1, 5, 16]. In the case of solvable groups, one can obtain exponent $\omega/2$ [2, 6].

In this article, we devise a more general way to circumvent the bottleneck summation, which depends on the structure of the group rather than knowledge of the representation theory. Our new recursive step permits us to decompose G via *two* subgroups H and K and recurse on H and K. See Theorem 3.7. One side-effect is an alternative proof of the $\omega/2$ exponent for solvable groups that does not require knowledge of the representation theory of the group (in Section 4). Our

[1] Note that exercise 13.16 in [3] claims that the exponent 1.5 can be reduced to 1.44 but this seems to be an error, as discussed in Section 2.

Group G	Upper bound	Reference
$\mathrm{SL}_2(\mathbb{F}_q)$	$\widetilde{O}(q\lvert G\rvert)$	Theorem 1.1 in [10]
$\mathrm{GL}_n(\mathbb{F}_q)$	$\widetilde{O}(q^n\lvert G\rvert)$	Theorem 4.3 in [15]
$\mathrm{PSp}_{2n}(\mathbb{F}_q)$	$\widetilde{O}(q^{5n-3}\lvert G\rvert)$	Theorem 5.14 in [13]
$O_{2n+1}(\mathbb{F}_q)$	$\widetilde{O}(q^{5n-3}\lvert G\rvert)$	Theorem 5.14 in [13]
$O^+_{2n}(\mathbb{F}_q)$, $n \geq 4$	$\widetilde{O}(q^{5n-6}\lvert G\rvert)$	Theorem 5.14 in [13]

Fig. 1. Previously best known running times for the generalized DFT over various families of linear groups. In this table, the $\widetilde{O}(\cdot)$ notation hides lower order terms and the dependence on n.

reduction bears some similarity to the double coset algorithm of Reference [18]; a key difference seems to be the use of fast matrix multiplication at an opportune time in the procedure.

1.1 Past and Related Work

A good description of past work in this area can be found in Section 13.5 of Reference [3]. The first algorithm generalizing beyond the Abelian case is from Beth in 1984 [2]; this algorithm is described in Section 2 in a form often credited jointly to Beth and Clausen. This algorithm was the best known for the general case of an arbitrary finite group prior to this work. Two other milestones are the $O(|G| \log |G|)$ algorithm for supersolvable groups from Baum [1], and the $O(|G| \log^3 |G|)$ algorithm for the symmetric group from Clausen [5]. The latter algorithm was improved to $O(|G| \log^2 |G|)$ by Maslen [16] and very recently to *linear* for the special case of S_{n-k}-invariant functions on S_n with $n > 2k$ [7]. Wreath products were studied by Rockmore [20], who obtained exponent one algorithms in certain cases.

In the 1990s, Maslen, Rockmore and coauthors developed the "separation of variables" approach [13], which relies on non-trivial decompositions along chains of subgroups via *Bratteli diagrams* and (again) detailed knowledge of the representation theory of the underlying groups. There is a rather large body of literature on this approach, and it has been applied to a wide variety of group algebras and more general algebraic objects. For a fuller description of this approach and the results obtained, the reader is referred to the surveys [17, 21], and the most recent article in this line of work [14].

For the present article, important results for comparison are the previous best-known results for linear groups of various sorts. We gather them in Figure 1. Notice that for each fixed dimension n, these all represent exponent α algorithms for $\alpha > 1$. Our methods give exponent $\omega/2$ algorithms for all of these groups, which translates to (the optimal) exponent one if $\omega = 2$. Using the current best upper bounds on ω, our methods give concrete improvements in small dimension in all cases; we explicitly highlight only the case of $\mathrm{SL}_2(\mathbb{F}_q)$ in this article.

1.2 Notation and Preliminaries

Throughout this article, we will use the phrase

$$\text{"}G \text{ has a generalized DFT using } O(|G|^{\alpha+\epsilon}) \text{ operations, for all } \epsilon > 0\text{,"}$$

where G is a finite group and $\alpha \geq 1$ is a real number. We mean by this that there are *universal* constants c_ϵ independent of the group G under consideration, so that for each $\epsilon > 0$, the operation count is at most $c_\epsilon |G|^{\alpha+\epsilon}$. Such an algorithm will be referred to as an "exponent α" algorithm. This comports with the precise definition of the exponent of matrix multiplication, ω: That there are universal constants b_ϵ for which $n \times n$ matrix multiplication can be performed using at most $b_\epsilon n^{\omega+\epsilon}$ operations, for each $\epsilon > 0$. Indeed, we will often report our algorithms' operation counts in terms of ω. In such cases, matrix multiplication is always used as a black box, so, for example, an

operation count of $O(|G|^{\omega/2})$ should be interpreted to mean: If one uses a fast matrix multiplication algorithm with exponent α (which may range from 2 to 3), then the operation count is $O(|G|^{\alpha/2})$. In particular, in real implementations, one might well use standard matrix multiplication and plug in 3 for ω in the operation count bound.

All logarithms are base 2. We use $\mathrm{Irr}(G)$ to denote the complete set of irreducible representations of G being used for the DFT at hand. In the presentation to follow, we assume the underlying field is $\mathbb{C}$; however, our algorithms work over any field $\mathbb{F}_{p^k}$ whose characteristic p does not divide the order of the group and for which k is sufficiently large for $\mathbb{F}_{p^k}$ to represent a complete set of irreducibles.

A basic fact is that $\sum_{\rho \in \mathrm{Irr}(G)} \dim(\rho)^2 = |G|$, which implies that for all $\rho \in \mathrm{Irr}(G)$, we have $\dim(\rho) \leq |G|^{1/2}$. An inequality that we use repeatedly is this one:

Proposition 1.1. *For any real number $\alpha > 2$, we have*

$$\sum_{\rho \in Irr(G)} \dim(\rho)^\alpha \leq |G|^{\alpha/2}.$$

Proof. Set $\rho_{\max}$ to be an irrep of largest dimension. We have

$$\sum_{\rho \in \mathrm{Irr}(G)} \dim(\rho)^\alpha \leq \dim(\rho_{\max})^{\alpha-2} \sum_{\rho \in \mathrm{Irr}(G)} \dim(\rho)^2 = \dim(\rho_{\max})^{\alpha-2}|G| \leq |G|^{\alpha/2},$$

where the last inequality used the fact that $\dim(\rho_{\max}) \leq |G|^{1/2}$. □

We also need Lev's Theorem:

Theorem 1.2 ([12]). *Every finite group G has a proper subgroup H of order at least $|G|^{1/2}$, unless G is cyclic of prime order.*

This is easily seen to be tight by considering the cyclic group of order p^2, for p prime.

In a few key places, we utilize the Kronecker product (or tensor product) of two matrices A and B, and there our convention is to name the indices of $A \otimes B$ so that

$$(A \otimes B)[(i, i'), (j, j')] = A[i, j]B[i', j'].$$

2 THE SINGLE SUBGROUP REDUCTION

In this section, we describe the recursive generalized DFT attributed to Beth and Clausen (see Reference [3]). Given a subgroup H of a finite group G, this reduction computes a DFT with respect to G via DFTs with respect to H. Our presentation makes use of fast matrix multiplication where possible, and so the running time will be expressed in terms of ω. A key definition is that of an *H-adapted basis* for the irreps of G. This is a basis in which the restriction of each irrep of G to H respects the direct sum decomposition into irreps of H. In concrete terms, this implies that for each irrep $\rho \in \mathrm{Irr}(G)$, while for general $g \in G$, $\rho(g)$ is a $\dim(\rho) \times \dim(\rho)$ matrix, for $g \in H$, $\rho(g)$ is a block-diagonal matrix with block sizes coming from the set $\{\dim(\sigma) : \sigma \in \mathrm{Irr}(H)\}$.

Theorem 2.1 (see Reference [3]). *Let G be a finite group and let H be a subgroup. Then we can compute a DFT with respect to G and an H-adapted basis, at a cost of $[G : H]$ many H-DFTs plus*

$$[G : H]|G| + [G : H]^2 \sum_{\sigma \in Irr(H)} O(\dim(\sigma)^{\omega+\epsilon})$$

operations, for all $\epsilon > 0$.

Proof. Let $g_1, g_2, \ldots, g_t$ be a system of distinct right coset representatives of H in G, so $t = [G : H]$. Let c be an element of $\mathbb{C}[G]$. We can write

$$c = \sum_{g \in G} c_g g = \sum_{i=1}^{t} \left(\sum_{h \in H} c_h^{(i)} h \right) g_i$$

for some elements $c^{(i)} = (\sum_{h \in H} c_h^{(i)} h) \in \mathbb{C}[H]$. By computing an H-DFT for each i, we obtain the elements

$$s_i = \sum_{h \in H} c_h^{(i)} \bigoplus_{\sigma \in \mathrm{Irr}(H)} \sigma(h).$$

Let $\overline{s_i}$ be the lift of s_i in which we repeat each $\sigma \in \mathrm{Irr}(H)$ as many times as it occurs in the irreps of G. We notice that

$$\sum_{g \in G} c_g \bigoplus_{\rho \in \mathrm{Irr}(G)} \rho(g) = \sum_{i=1}^{t} \overline{s_i} \cdot \left(\bigoplus_{\rho \in \mathrm{Irr}(G)} \rho(g_i) \right).$$

Moreover, since we are using an H-adapted basis, each of the t matrix multiplications is the product of a block-diagonal matrix having blocks whose dimensions are those of the irreps of H, with a block diagonal matrix having blocks whose dimensions are those of the irreps of G. If $n_{\sigma,\rho}$ denotes the number of occurrences of $\sigma \in \mathrm{Irr}(H)$ in $\rho \in \mathrm{Irr}(G)$, then the cost of performing this structured matrix multiplication is at most

$$\begin{aligned}
\sum_{\sigma \in \mathrm{Irr}(H)} \sum_{\rho \in \mathrm{Irr}(G)} n_{\sigma,\rho} O(\dim(\sigma)^{\omega+\epsilon}) \left\lceil \frac{\dim(\rho)}{\dim(\sigma)} \right\rceil &= \sum_{\sigma \in \mathrm{Irr}(H)} O(\dim(\sigma)^{\omega-1+\epsilon}) \sum_{\rho \in \mathrm{Irr}(G)} n_{\sigma,\rho} \dim(\rho) \\
&= \sum_{\sigma \in \mathrm{Irr}(H)} O(\dim(\sigma)^{\omega-1+\epsilon}) \dim(\sigma)[G : H] \\
&= \sum_{\sigma \in \mathrm{Irr}(H)} O(\dim(\sigma)^{\omega+\epsilon})[G : H],
\end{aligned}$$

where the second-to-last equality used Frobenius reciprocity: $n_{\sigma,\rho}$ also equals the number of times ρ occurs in the induction of σ from H up to G, and then $\sum_\rho n_{\sigma,\rho} \dim(\rho)$ is easily seen to be the dimension of the induced representation, which is $\dim(\sigma)[G : H]$. We have to do $[G : H]$ many of these structured multiplications, and then sum them up. The summing costs $[G : H]|G|$ many operations, since the block-diagonal matrices we are summing have, in general, $|G|$ nonzeros. □

We note that this final sum, which costs $|G|[G : H]$ operations, cannot be accelerated by fast matrix multiplication, and this appears to have been overlooked in the claim in Referencee [3] that by using fast matrix multiplication together with Theorem 1.2 one can achieve an upper bound of $O(|G|^{1.44})$ for all finite groups G. Indeed, when $|H| = |G|^{1/2}$, which may be in the worst case, the $|G|[G : H]$ term by itself is at least $|G|^{3/2}$. Our "double subgroup reduction" can be seen as a means to avoid having to directly compute this bottleneck sum.

At the expense of a slightly coarser upper bound, we can remove the requirement of an H-adapted basis, which will simplify our use of Theorem 2.1 in recursive algorithms later.

Corollary 2.2. *Let G be a finite group and let H be a subgroup. Then we can compute a DFT with respect to G at a cost of $[G : H]$ many H-DFTs plus $O([G : H]^2 |H|^{\omega/2+\epsilon})$ operations for all sufficiently small $\epsilon > 0$.*

Proof. Using Proposition 1.1 with $\alpha = \omega + \epsilon$, the cost from the statement of Theorem 2.1 can be upper bounded by

$$[G : H]|G| + [G : H]^2 \sum_{\sigma \in \mathrm{Irr}(H)} O(\dim(\sigma)^{\omega+\epsilon}) \leq O([G : H]^2 |H|^{\omega/2+\epsilon}). \tag{1}$$

Note that in Theorem 2.1, the DFT is with respect to an H-adapted basis. At a cost of

$$\sum_{\rho \in \mathrm{Irr}(G)} O(\dim(\rho)^{\omega+\epsilon}) \leq O(|G|^{\omega/2+\epsilon}) \tag{2}$$

operations (again using Proposition 1.1 with $\alpha = \omega + \epsilon$), we can change an arbitrary basis to an H-adapted basis, to which we apply Theorem 2.1 and then change back to the original basis. Both expression (1) and expression (2) are upper bounded by $O([G : H]^2 |H|^{\omega/2+\epsilon})$, provided $\omega + 2\epsilon \leq 4$. □

The single-subgroup reduction works best when the subgroup H is large. Lev's Theorem (Theorem 1.2) guarantees a subgroup of size at least $|G|^{1/2}$. Using this, one obtains the following recursive algorithm, whose bound, using only that $\omega \leq 3$, matches Theorem 13.48 in the presentation in Reference [3].

Theorem 2.3. *For every finite group G, there is an exponent one $+\omega/4$ algorithm computing the DFT with respect to G.*

Proof. Fix G. We apply Corollary 2.2 recursively.

If G is a p-group, then we apply Theorem 4.2 (actually, we only need to do this when G is cyclic of prime order). If G is the trivial group, then the DFT is trivial as well. Otherwise, according to Theorem 1.2, there is a subgroup H of size at least $|G|^{1/2}$, to which we apply Corollary 2.2.

Set $\delta = \min\{\epsilon, 0.1\}$, and give names to some constants:

- Let B_δ be the constant hidden in the $[G : H]^2 \cdot O(|H|^{\omega/2+\delta})$ notation of Corollary 2.2.
- Let B be the constant hidden in the $O(|G| \log |G|)$ notation of Theorem 4.2.

Let $T(n)$ denote an upper bound on the operation count of this recursive algorithm for any group G of order n. For each fixed $\epsilon > 0$, we will prove by induction on n that, for a universal constant C_ϵ,

$$T(n) \leq C_\epsilon n^{1+\frac{\omega}{4}+\epsilon} \log^2 n.$$

This clearly holds for the base case of a p-group or the trivial group, provided $C_\epsilon > B$.

When we apply Corollary 2.2 recursively, the cost is at most

$$[G : H] \cdot T(|H|) + [G : H]^2 \cdot B_\delta |H|^{\omega/2+\delta},$$

where $|H| \geq |G|^{1/2}$. If we set γ such that $|H| = |G|^\gamma$, and thus $1/2 \leq \gamma \leq 1$, and apply the induction hypothesis, then we obtain

$$\begin{aligned} T(n) &\leq C_\epsilon n^{1-\gamma} n^{\gamma(1+\frac{\omega}{4}+\epsilon)} \log^2(n/2) + B_\delta n^{2(1-\gamma)} n^{\gamma(\omega/2+\delta)} \\ &< C_\epsilon n^{1+\omega/4+\epsilon} (\log n)(\log n - 1) + B_\delta n^{1+\frac{\omega}{4}+\frac{\delta}{2}}, \end{aligned}$$

which is at most $C_\epsilon n^{1+\frac{\omega}{4}+\epsilon} \log^2 n$ as long as $C_\epsilon \geq B_\delta$. □

3 THE DOUBLE SUBGROUP REDUCTION

This section contains our main algorithmic result. Given two subgroups H, K of a finite group G, we show how to compute a DFT with respect to G, via DFTs with respect to H and K. We first show how to obtain an intermediate representation in terms of tensor products of the irreps of H and the irreps of K:

Lemma 3.1. *Let H and K be subgroups of G and let c be an element of $\mathbb{C}[G]$ supported on HK. Fix a way of writing $g = hk$ for each $g \in HK$ (this is unique iff $H \cap K = \{1\}$). We can compute*

$$\sum_{g=hk \in HK} c_g \bigoplus_{\sigma \in Irr(H), \tau \in Irr(K)} \sigma(h) \otimes \tau(k)$$

by performing $|H|$ many K-DFTs and $|K|$ many H-DFTs.

Proof. Using the chosen way of writing $g = hk$, we can write

$$c = \sum_{g=hk \in HK} c_g g = \sum_{h \in H} h \cdot \left(\sum_{k \in K} c_k^{(h)} k \right)$$

for some elements $c^{(h)} = (\sum_{k \in K} c_k^{(h)} k) \in \mathbb{C}[K]$. Specifically, among all h, k pairs such that $hk = g$, we take $c_k^{(h)}$ equal to c_g for the chosen h, k pair, and zero for the other pairs. We perform $|H|$ many K-DFTs to compute for each $h \in H$:

$$s_h = \sum_{k \in K} c_k^{(h)} \bigoplus_{\tau \in \mathrm{Irr}(K)} \tau(k).$$

We use the notation $s_h[\tau, u, v]$ to refer to entry (u, v) of component τ in the direct sum. Then we perform $|K|$ many H-DFTs to compute for each $\tau \in \mathrm{Irr}(K)$ and $u, v \in [\dim(\tau)]$,

$$t_{\tau,u,v} = \sum_{h \in H} s_h[\tau, u, v] \bigoplus_{\sigma \in \mathrm{Irr}(H)} \sigma(h).$$

Note that $t_{\tau,u,v}[\sigma, x, y]$ is the $((x, u), (y, v))$ entry of $\sum_{h \in H, k \in K} c_k^{(h)} \sigma(h) \otimes \tau(k)$ and then using our choice of $c^{(h)}$, we find that we have computed:

$$\sum_{g=hk \in HK} c_{hk} \bigoplus_{\sigma \in \mathrm{Irr}(H), \tau \in \mathrm{Irr}(K)} \sigma(h) \otimes \tau(k)$$

as promised. □

The following is an important (and known) general observation (see, e.g., Lemma 4.3.1 in Reference [8]):

Lemma 3.2. *If A is an $n_1 \times n_2$ matrix, B is an $n_2 \times n_3$ matrix, and C is an $n_3 \times n_4$ matrix, then the product ABC can be computed by multiplying $A \otimes C^T$ (which is an $n_1 n_4 \times n_2 n_3$ matrix) by B viewed as an $n_2 n_3$-vector.*

Proof. Observe that

$$(ABC)[i_1, i_4] = \sum_{i_2, i_3} A[i_1, i_2] B[i_2, i_3] C[i_3, i_4]$$

and

$$((A \otimes C^T) \cdot B)[(i_1, i_4)] = \sum_{i_2, i_3} (A \otimes C^T)[(i_1, i_4), (i_2, i_3)] B[(i_2, i_3)] = \sum_{i_2, i_3} A[i_1, i_2] C[i_3, i_4] B[i_2, i_3]. \quad \square$$

This $n_1n_4 \times n_2n_3$-matrix-vector multiplication costs $O(n_1n_4n_2n_3)$ operations. More importantly, we have the following:

Corollary 3.3. *If A and C are as in Lemma 3.2, and square (so $n_1 = n_2$ and $n_3 = n_4$), and we have several $n_2 \times n_3$ matrices, $B_1, B_2, \ldots, B_\ell$, then we can compute AB_iC for all i using $A \otimes C^T$, at a cost of*

$$O((n_2n_3)^{\omega-1+\epsilon} \cdot \max\{n_2n_3, \ell\})$$

operations, for all $\epsilon > 0$.

Proof. Set $N = n_1n_4 = n_2n_3$. If $\ell \le N$, then this can be accomplished with a single $N \times N$ matrix multiplication, at a cost of $O(N^{\omega+\epsilon})$ operations, by the definition of ω. If $\ell > N$, then this can be accomplished with $\lceil \ell/N \rceil$ many $N \times N$ matrix multiplications, at a cost of $O(\ell \cdot N^{\omega-1+\epsilon})$ operations. □

Now we show how to lift from the intermediate representation to the space of irreducibles of G. We need some notation. For $\sigma \in \mathrm{Irr}(H), \tau \in \mathrm{Irr}(K), \rho \in \mathrm{Irr}(G)$, let $n_{\sigma,\rho}$ be the number of occurrences of σ in the restriction of ρ to H, and let $m_{\tau,\rho}$ be the number of occurrences of τ in the restriction of ρ to K.

Lemma 3.4. *There is a linear map*

$$\phi_{G,H,K} : \prod_{\sigma \in Irr(H), \tau \in Irr(K)} \mathbb{C}^{(\dim(\sigma)\dim(\tau))^2} \to \prod_{\rho \in Irr(G)} \mathbb{C}^{\dim(\rho)^2}$$

that maps $\bigoplus_{\sigma \in Irr(H), \tau \in Irr(K)} \sigma(h) \otimes \tau(k)$ to $\bigoplus_{\rho \in Irr(G)} \rho(hk)$ for all $h \in H, k \in K$. The map $\phi_{G,H,K}$ can be computed using

$$\sum_{\sigma \in Irr(H), \tau \in Irr(K)} O\left((\dim(\sigma)\dim(\tau))^{\omega-1+\epsilon} \cdot \max\left\{\dim(\sigma)\dim(\tau), \sum_{\rho \in IrrG} n_{\sigma,\rho}m_{\tau,\rho}\right\}\right) + \sum_{\rho \in Irr(G)} O(\dim(\rho)^{\omega+\epsilon})$$

operations, for all $\epsilon > 0$.

Proof. Let $\mathrm{Irr}^*(H)$ be the multiset of irreducibles of H in the multiplicities that they occur in the restrictions to H of $\mathrm{Irr}(G)$, and let $\mathrm{Irr}^*(K)$ be the multiset of irreducibles of K in the multiplicities that they occur in the restrictions to K of $\mathrm{Irr}(G)$. Let S be the change of basis matrix taking $\oplus_{\sigma \in \mathrm{Irr}^*(H)}\sigma$ to $\oplus_{\rho \in \mathrm{Irr}(G)}\rho$ and let T be the change of basis matrix taking $\oplus_{\tau \in \mathrm{Irr}^*(K)}\tau$ to $\oplus_{\rho \in \mathrm{Irr}(G)}\rho$. Then for each $h \in H, k \in K$, we have

$$S\left(\bigoplus_{\sigma \in \mathrm{Irr}^*(H)} \sigma(h)\right)S^{-1}T\left(\bigoplus_{\tau \in \mathrm{Irr}^*(K)} \tau(k)\right)T^{-1} = \bigoplus_{\rho \in \mathrm{Irr}(G)} \rho(hk).$$

Set $M = S^{-1}T$, and consider the expression

$$\left(\bigoplus_{\sigma \in \mathrm{Irr}^*(H)} \sigma(h)\right)M\left(\bigoplus_{\tau \in \mathrm{Irr}^*(K)} \tau(k)\right). \tag{3}$$

Note that both M and the above product are block-diagonal matrices with blocks of dimension $\dim(\rho)$ as ρ runs through $\mathrm{Irr}(G)$. Now, for each $\rho \in \mathrm{Irr}(G)$, a given $\sigma \in \mathrm{Irr}(H)$ occurs $n_{\sigma,\rho}$ times and

a given $\tau \in \mathrm{Irr}(K)$ occurs $m_{\tau,\rho}$ times; therefore, we are computing $\sigma(h)B_i\tau(k)$ for p distinct sub-matrices B_i of M, where $p = \sum_{\rho\in\mathrm{Irr}(G)} n_{\sigma,\rho}m_{\tau,\rho}$. By Corollary 3.3, each such a batch can be computed by taking a product of $\sigma(h) \otimes \tau(k)^T$ with a matrix whose columns are the B_i sub-matrices, viewed as vectors. This is linear in the entries of $\sigma(h) \otimes \tau(k)$ and costs

$$O\left((\dim(\sigma)\dim(\tau))^{\omega-1+\epsilon} \cdot \max\left\{\dim(\sigma)\dim(\tau), \sum_{\rho\in\mathrm{Irr}(G)} n_{\sigma,\rho}m_{\tau,\rho}\right\}\right)$$

operations. Finally, we need to multiply Equation (3) by S on the left and T^{-1} on the right; both are block-diagonal matrix multiplications that cost $\sum_{\rho\in\mathrm{Irr}(G)} O(\dim(\rho)^{\omega+\epsilon})$ operations.

Note that Equation (3) specifies a matrix multiplication problem with a format and a pattern of repeated blocks that is *independent* of h and k (it depends only on G, H, K). The just-described map is therefore the same for each h, k, and we call it $\phi_{G,H,K}$. Both the applications of Corollary 3.3 and the pre- and post- multiplication by S and T^{-1} are linear in the entries of $\bigoplus_{\sigma\in\mathrm{Irr}(H),\tau\in\mathrm{Irr}(K)} \sigma(h) \otimes \tau(k)$, as required. □

Now we use elementary facts from representation theory to bound the complexity estimate in Lemma 3.4 in terms of $|H|, |K|, |G|$.

Lemma 3.5. *For all finite groups G and subgroups H, K, the expression*

$$\sum_{\sigma\in Irr(H),\tau\in Irr(K)} O\left((\dim(\sigma)\dim(\tau))^{\omega-1+\epsilon} \cdot \max\left\{\dim(\sigma)\dim(\tau), \sum_{\rho\in IrrG} n_{\sigma,\rho}m_{\tau,\rho}\right\}\right) + \sum_{\rho\in Irr(G)} O(\dim(\rho)^{\omega+\epsilon})$$

is upper bounded by $O((|H||K|)^{\omega/2+\epsilon/2} + |G|^{\omega/2+\epsilon/2})$.

Proof. We use only the fact that for each $\rho \in \mathrm{Irr}(G)$,

$$\sum_{\sigma\in\mathrm{Irr}(H)} \dim(\sigma)n_{\sigma,\rho} = \dim(\rho), \tag{4}$$

and similarly

$$\sum_{\tau\in\mathrm{Irr}(K)} \dim(\tau)m_{\tau,\rho} = \dim(\rho), \tag{5}$$

together with the fact that the sum of the squares of the dimensions of the irreps of a group is the order of that group (which implies that the maximum dimension is at most the square root of the order of the group).

We observe that by replacing the "max" with addition,

$$\sum_{\sigma\in\mathrm{Irr}(H),\tau\in\mathrm{Irr}(K)} O\left((\dim(\sigma)\dim(\tau))^{\omega-1+\epsilon} \cdot \max\left\{\dim(\sigma)\dim(\tau), \sum_{\rho\in IrrG} n_{\sigma,\rho}m_{\tau,\rho}\right\}\right)$$
$$\leq \sum_{\sigma\in\mathrm{Irr}(H),\tau\in\mathrm{Irr}(K)} O\left((\dim(\sigma)\dim(\tau))^{\omega-1+\epsilon} \cdot \left(\dim(\sigma)\dim(\tau) + \sum_{\rho\in IrrG} n_{\sigma,\rho}m_{\tau,\rho}\right)\right)$$

We know that

$$\sum_{\sigma\in\mathrm{Irr}(H),\,\tau\in\mathrm{Irr}(K)} (\dim(\sigma)\dim(\tau))^{\omega-1+\epsilon}\cdot\dim(\sigma)\dim(\tau)$$
$$=\left(\sum_{\sigma\in\mathrm{Irr}(H)}\dim(\sigma)^{\omega+\epsilon}\right)\cdot\left(\sum_{\tau\in\mathrm{Irr}(K)}\dim(\tau)^{\omega+\epsilon}\right)\le(|H||K|)^{\omega/2+\epsilon/2},$$

where the last inequality applied Proposition 1.1 twice, with $\alpha=\omega+\epsilon$. Also, we know that

$$\sum_{\sigma\in\mathrm{Irr}(H),\,\tau\in\mathrm{Irr}(K)} (\dim(\sigma)\dim(\tau))^{\omega-1+\epsilon}\cdot\left(\sum_{\rho\in\mathrm{Irr}G} n_{\sigma,\rho}m_{\tau,\rho}\right)$$
$$=\sum_{\rho\in\mathrm{Irr}(G)}\left(\sum_{\sigma\in\mathrm{Irr}(H)}\dim(\sigma)^{\omega-1+\epsilon}n_{\sigma,\rho}\right)\cdot\left(\sum_{\tau\in\mathrm{Irr}(K)}\dim(\tau)^{\omega-1+\epsilon}m_{\tau,\rho}\right)$$
$$\le\sum_{\rho\in\mathrm{Irr}(G)}\left(|H|^{(\omega-2+\epsilon)/2}\cdot\sum_{\sigma\in\mathrm{Irr}(H)}\dim(\sigma)n_{\sigma,\rho}\right)\cdot\left(|K|^{(\omega-2+\epsilon)/2}\cdot\sum_{\tau\in\mathrm{Irr}(K)}\dim(\tau)m_{\tau,\rho}\right)$$
$$=\sum_{\rho\in\mathrm{Irr}(G)}|H|^{(\omega-2+\epsilon)/2}|K|^{(\omega-2+\epsilon)/2}\dim(\rho)^2=(|H||K|)^{(\omega-2+\epsilon)/2}|G|,$$

where the second-to-last equality used Equations (4) and (5). If $|H||K|\le|G|$, then this expression is at most $|G|^{\omega/2+\epsilon/2}$; if $|H||K|>|G|$, then this expression is at most $(|H||K|)^{\omega/2+\epsilon/2}$. Finally, we have that the final term in the main expression, $\sum_{\rho\in\mathrm{Irr}(G)}O(\dim(\rho)^{\omega+\epsilon})$, is at most $O(|G|^{\omega/2+\epsilon/2})$, by Proposition 1.1 with $\alpha=\omega+\epsilon$, and the lemma follows. □

Our main theorems put everything together:

Theorem 3.6. *Let G be a finite group, let H,K be subgroups, and let $x\in G$ be any element. Fix a way of writing $g=hk$ for each $g\in HK$ (this is unique iff $H\cap K=\{1\}$). Let $c\in\mathbb{C}[G]$ be supported on HKx. Then we can compute*

$$\sum_{g=hkx\in HKx} c_g\cdot\bigoplus_{\rho\in Irr(G)}\rho(g)$$

at the cost of $|H|$ many K-DFTS, $|K|$ many H-DFTs, plus $O(|G|^{\omega/2+\epsilon}+(|H||K|)^{\omega/2+\epsilon})$ operations for all $\epsilon>0$.

Proof. Set $c'_g=c_{gx}$ and notice that c' is supported on HK. Apply Lemma 3.1 on c' to compute

$$\sum_{g=hk\in HK} c'_g\bigoplus_{\sigma\in\mathrm{Irr}(H),\,\tau\in\mathrm{Irr}(K)}\sigma(h)\otimes\tau(k).$$

Next, apply the linear map $\phi_{G,H,K}$ to obtain (by linearity) $\sum_{g=hk\in HK}c'_g\bigoplus_{\rho\in\mathrm{Irr}(G)}\rho(hk)$, and, finally, multiply by $\oplus_{\rho\in\mathrm{Irr}(G)}\rho(x)$ on the right, at a cost of $\sum_{\rho\in\mathrm{Irr}(G)}\dim(\rho)^2\le O(|G|^{\omega/2+\epsilon})$ operations (by Proposition 1.1 with $\alpha=\omega+\epsilon$). The result is

$$\sum_{g=hk\in HK} c'_g\bigoplus_{\rho\in\mathrm{Irr}(G)}\rho(gx)=\sum_{g'\in HKx} c_{g'}\bigoplus_{\rho\in\mathrm{Irr}(G)}\rho(g'),$$

as promised. □

By translating HK around, we cover all of G, leading to our main theorem:

Theorem 3.7 (main). *Let G be a finite group and let H, K be subgroups. Then we can compute the DFT with respect to G at the cost of $|H|$ many K-DFTS, $|K|$ many H-DFTs, plus $O(|G|^{\omega/2+\epsilon} + (|H||K|)^{\omega/2+\epsilon})$ operations, all repeated $r = O(\frac{|G|\ln(|G|)}{|HK|})$ many times, for all $\epsilon > 0$. If $G = HK$, then we may take $r = 1$.*

Proof. We argue that there exist $x_1, x_2, \ldots, x_r \in G$ so that $\cup_i HKx_i = G$. Then a G-DFT can be computed by applying Theorem 3.6 r times with these translations. The existence of the x_i is a standard application of the probablistic method: For randomly chosen x_i, the probability $\cup_i HKx_i$ fails to contain a given $g \in G$ is $(1 - |HK|/|G|)^r$, and the r specified in the theorem statement makes this quantity strictly less than $1/|G|$, so a union bound finishes the argument. □

4 EXPONENT $\omega/2$ FOR FINITE SOLVABLE GROUPS

We show how to derive algorithms for all solvable groups via our reduction, matching the exponent $\omega/2$ algorithm of References [2, 6]. An advantage of our approach is that we do not need to rely on knowledge of the representation theory of G.

We begin with a key definition:

Definition 4.1. A finite group G is *supersolvable* if there is a sequence of subgroups

$$\{1\} = G_0 \lhd G_1 \lhd G_2 \lhd \cdots \lhd G_k = G,$$

such that each G_i is normal in G, and for all i, G_i/G_{i-1} is cyclic of prime order.

A *solvable* finite group G is one in which the requirement that each G_i is normal in G (rather than just G_{i+1}) is removed. An early result in the area of fast generalized DFTs was Baum's algorithm, which gives a fast DFT for all *supersolvable* groups.

Theorem 4.2 (Baum). *There is an algorithm that uses $O(|G| \log |G|)$ operations to compute the generalized DFT over G if G is supersolvable.*

An important class of supersolvable groups are p-groups. Together with this fact, the result of the previous section makes it quite easy to obtain an algorithm for all solvable groups. We need the following classical result of Hall:

Theorem 4.3 (Hall). *Let G be a finite solvable group of order ab, with $(a, b) = 1$. Then there exists a subgroup $H \subseteq G$ of order a.*

From this we obtain the following:

Theorem 4.4. *Let G be a finite solvable group. Then a G-DFT can be computed in $O(|G|^{\omega/2+\epsilon})$ operations for all $\epsilon > 0$.*

Proof. Take $\delta = \epsilon/2$. Let $A_\delta \geq 1$ be the constant hidden in the $O(|G|^{\omega/2+\delta} + (|H||K|)^{\omega/2+\delta})$ notation in Theorem 3.7. Let B be the constant in the $O(|G| \log |G|)$ expression in the statement of Theorem 4.2. It suffices to prove that for any finite group G with $|G|$ having k distinct prime factors, a G-DFT can be computed in

$$(4A_\delta)^{\log k} |G|^{\omega/2+\delta} B \log |G|$$

operations, because for sufficiently large G, we have

$$(4A_\delta)^{\log k} B \log |G| \leq (4A_\delta)^{\log \log |G|} B \log |G| \leq |G|^\delta.$$

The proof is by induction on the number of distinct prime factors in the order of G. For the base case of $k = 1$, G is a p-group, hence supersolvable, and we apply Theorem 4.2.

Now, suppose $|G| = p_1^{a_1} \dots p_k^{a_k}$, where $p_1, \dots, p_k$ are distinct primes, and then $|G| = ab$, where a and b each has no more than $k/2$ distinct prime factors and $(a, b) = 1$. Applying Hall's theorem (twice) there are subgroups H, K of order a and b, respectively. Since $(a, b) = 1$, we must have $H \cap K = \{1\}$, and then $G = HK$, because $|G| = ab$.

We can then apply Theorem 3.7 to reduce to the case of computing $|H|$ many K-DFTs and K many H-DFTs, at a cost of $2A_\delta |G|^{\omega/2+\delta}$ operations. But H and K are both solvable, and hence by the induction hypothesis, these two sets of DFTs cost at most

$$\begin{aligned} &|H| \cdot (4A_\delta)^{\log(k/2)} |K|^{\omega/2+\delta} B \log |K| + |K| \cdot (4A_\delta)^{\log(k/2)} |H|^{\omega/2+\delta} B \log |H| \\ &\leq \frac{2}{4A_\delta} (4A_\delta)^{\log k} |G|^{\omega/2+\delta} B \log |G| \end{aligned}$$

operations. Together with the $2A_\delta |G|^{\omega/2+\delta}$ overhead, this is no more than

$$(4A_\delta)^{\log k} |G|^{\omega/2+\delta} B \log |G|$$

operations, as required. □

5 EXPONENT $\omega/2$ FOR FINITE GROUPS OF LIE TYPE

One of the main payoffs of Theorem 3.7 is exponent $\omega/2$ algorithms for finite groups of Lie type. This is because groups of Lie type have an "*LDU*-type" decomposition that is well suited to Theorem 3.7. We describe these decompositions and the resulting DFT algorithms in this section. All of our "*LDU*-type" decompositions of groups of Lie type into three subgroups give rise to the following DFT algorithm:

Theorem 5.1. *Let H_1, H_2, H_3 be subgroups of group G, and suppose all three are either p-groups or Abelian. Moreover, suppose that H_1H_2 is a subgroup of G and that $H_1 \cap H_2 = \{1\}$ and $H_1H_2 \cap H_3 = \{1\}$. Then there is a generalized DFT for G that uses at most*

$$O\left(|G|^{\omega/2+\epsilon} \frac{|G| \log |G|}{|H_1||H_2||H_3|}\right)$$

operations for all $\epsilon > 0$.

Proof. We apply Theorem 3.7 to the pair H_1H_2 and H_3 at a cost of $O(|G|^{\omega/2+\epsilon})$ operations plus $|H_1H_2|$ many H_3-DFTs and $|H_3|$ many H_1H_2-DFTs. This is all repeated,

$$r = O\left(\frac{|G| \log |G|}{|H_1||H_2||H_3|}\right),$$

many times. The H_3-DFTs cost $O(|H_3| \log |H_3|)$ operations, because H_3 is Abelian or a p-group (via Theorem 4.2). We apply Theorem 3.7 once more to H_1, H_2, at a cost of $O(|H_1H_2|^{\omega/2+\epsilon})$ operations plus $|H_1|$ many H_2-DFTs and $|H_2|$ many H_1-DFTs. Each H_1-DFT costs $O(|H_1| \log |H_1|)$ operations, because H_1 is Abelian or a p-group, and the same is true for each H_2-DFT. Altogether, the cost is

$$\begin{aligned} r \cdot \Big[O(|G|^{\omega/2+\epsilon}) \quad &+ \quad |H_1H_2| \cdot O(|H_3| \log |H_3|) \\ &+ \quad |H_3| \cdot \Big(O(|H_1H_2|^{\omega/2+\epsilon}) + |H_1| \cdot O(|H_2| \log |H_2|) + H_2 \cdot O(|H_1 \log H_1|) \Big) \Big] \end{aligned}$$

operations, which is as claimed. □

From Carter [4], we have that all finite simple groups of Lie type (except the Tits group) have a *split (B, N)-pair*, which implies the following structure:

$$G = \sqcup_{w \in W} B \overline{w} U_w,$$

Name	Family	$\lvert W\rvert$	$\lvert G\rvert$
Chevalley	$A_\ell(q)$	$(\ell+1)!$	$q^{\Theta(\ell^2)}$
	$B_\ell(q)$	$2^\ell \ell!$	$q^{\Theta(\ell^2)}$
	$C_\ell(q)$	$2^\ell \ell!$	$q^{\Theta(\ell^2)}$
	$D_\ell(q)$	$2^{\ell-1}\ell!$	$q^{\Theta(\ell^2)}$
Exceptional Chevalley	$E_6(q)$	$O(1)$	$q^{\Theta(1)}$
	$E_7(q)$	$O(1)$	$q^{\Theta(1)}$
	$E_8(q)$	$O(1)$	$q^{\Theta(1)}$
	$F_4(q)$	$O(1)$	$q^{\Theta(1)}$
	$G_2(q)$	$O(1)$	$q^{\Theta(1)}$
Steinberg	${}^2A_\ell(q^2)$	$2^{\lceil \ell/2\rceil}\lceil \ell/2\rceil!$	$q^{\Theta(\ell^2)}$
	${}^2D_\ell(q^2)$	$2^{\ell-1}(\ell-1)!$	$q^{\Theta(\ell^2)}$
	${}^2E_6(q^2)$	$O(1)$	$q^{\Theta(1)}$
	${}^3D_4(q^3)$	$O(1)$	$q^{\Theta(1)}$
Suzuki	${}^2B_2(q)$, $q = 2^{2n+1}$	$O(1)$	$q^{\Theta(1)}$
Ree	${}^2F_4(q)$, $q = 3^{2n+1}$	$O(1)$	$q^{\Theta(1)}$
	${}^2G_2(q)$, $q = 3^{2n+1}$	$O(1)$	$q^{\Theta(1)}$

Fig. 2. Families of finite groups G of Lie type, together with the size of their associated Weyl group W. These include all simple finite groups other than cyclic groups, the alternating groups, the 26 sporadic groups, and the Tits group. See References [12, 22] for sources.

where B and N are subgroups, W is the Weyl group (i.e. $W = B/(B \cap N)$), $B = UT$ with T a maximal torus (hence Abelian), and U, T are complements in B. The notation $\overline{w}$ denotes a lift of w from W to N. The U_w are subgroups of U, and U is a p-group. This decomposition is "with uniqueness of expression," which implies that $|B\overline{w}U_w| = |B||U_w|$ for each w.

From this description, we easily have the very general result:

Theorem 5.2. *Let G be a finite group with a split (B, N)-pair, with associated Weyl group W. Then there is a fast DFT over G that uses $O(|G|^{\omega/2+\epsilon}|W|)$ operations for all $\epsilon > 0$.*

Proof. Fix the w maximizing the size of the double coset $B\overline{w}U_w$, and note that $|BU_w^{\overline{w}}| = |B\overline{w}U_w| \geq |G|/|W|$ (where $U_w^{\overline{w}}$ is the conjugate subgroup $\overline{w}U_w\overline{w}^{-1}$). As noted, this size is $|B||U_w|$, and hence $B \cap U_w^{\overline{w}} = \{1\}$. Also from the description above, $B = UT$ with $U \cap T = \{1\}$; T is Abelian, and $U, U_w^{\overline{w}}$ are p-groups. We are then in the position to apply Theorem 5.1, which yields the claimed operation count. □

As one can see from Figure 2, for families of finite simple groups of Lie type, the Weyl group always has order that is $|G|^{o(1)}$, so this algorithm has exponent $\omega/2$, which is best-possible if $\omega = 2$. Next, we explicitly work out the more common cases of the general linear, orthogonal, and symplectic families, and their variants. The overhead coming from the parameter r in Theorem 3.7 in each case is somewhat smaller than the worst-case bound of $O(|W| \log |G|)$ coming from (the very general) Theorem 5.2; instead, it approaches $O(\log |G|)$ as the underlying field size q approaches infinity.

5.1 The Groups $\mathrm{GL}_n(\mathbb{F}_q)$ and $\mathrm{SL}_n(\mathbb{F}_q)$

The easiest example for applying Theorem 5.1 is the general linear group.

Theorem 5.3. *For each n and prime power q, there is a generalized DFT for the group $G = GL_n(\mathbb{F}_q)$ that uses $O(|G|^{\omega/2+\epsilon})$ operations for all $\epsilon > 0$.*

Proof. The three subgroups H_1, H_2, H_3 are the set of lower-triangular matrices with ones on the diagonal, the set of diagonal matrices, and the set of upper-triangular matrices with ones on the diagonal, which have sizes $q^{(n^2-n)/2}$, $(q-1)^n$, and $q^{(n^2-n)/2}$, respectively. In the notation of Theorem 5.1, we have

$$r = O\left(\frac{|G|\log|G|}{|H_1||H_2||H_3|}\right) \leq O\left(\frac{q}{q-1}\right)^n (n^2 \log q),$$

which can be absorbed into the $|G|^\epsilon$ term. □

For $\mathrm{SL}_n(\mathbb{F}_q)$ the only difference is that the diagonal matrices must have determinant one, so the size of that subgroup is $(q-1)^{n-1}$ instead of $(q-1)^n$; the group itself is also smaller by a factor of $q-1$. We obtain in exactly the same way as for Theorem 5.3:

Theorem 5.4. *For each n and prime power q, there is a generalized DFT for $G = SL_n(\mathbb{F}_q)$ that uses $O(|G|^{\omega/2+\epsilon})$ operations for all $\epsilon > 0$.*

Since the two-dimensional case has attracted a lot of attention, we record that result separately for concreteness as follows:

Theorem 5.5. *For each prime power q, there is a generalized DFT for $G = SL_2(\mathbb{F}_q)$ that uses $O(|G|^{\omega/2+\epsilon})$ operations for all $\epsilon > 0$.*

Proof. Let H_1 be the set of lower triangular matrices with ones on the diagonal, H_2 be the set of diagonal matrices with determinant 1, and H_3 be the set of upper triangular matrices with ones on the diagonal. These are all subgroups, each pairwise intersection is $\{1\}$, and we have H_1H_2 is a subgroup. All three subgroups are Abelian, with orders $q, q-1$, and q, respectively. Since $|G| = q^3 - q$, we have in this case that $|H_1H_2||H_3| = |G|$ and hence $H_1H_2H_3 = G$. We can perform the DFT by applying Theorem 3.7 to H_1H_2 and H_3 and then to H_1 and H_2. The overall cost is

$$\begin{aligned} O(|G|^{\omega/2+\epsilon}) \quad &+ \quad |H_1H_2| \cdot O(|H_3|\log|H_3|) \\ &+ \quad |H_3| \cdot \left(O(|H_1H_2|^{\omega/2+\epsilon}) + |H_1| \cdot O(|H_2|\log|H_2|) + H_2 \cdot O(|H_1\log H_1|)\right), \end{aligned}$$

which simplifies to the claimed operation count. □

5.2 The Symplectic Groups $\mathrm{Sp}_{2n}(\mathbb{F}_q)$

A symplectic group of dimension $2n$ over $\mathbb{F}_q$ is the subgroup of invertible matrices that preserve a symplectic form; all symplectic forms are equivalent under a change of basis, so concretely we may take $\mathrm{Sp}_{2n}(\mathbb{F}_q)$ to be the set of all matrices $A \in \mathrm{GL}_{2n}(\mathbb{F}_q)$ such that

$$A^T Q A = Q, \text{ where } Q = \begin{pmatrix} 0 & J \\ -J & 0 \end{pmatrix},$$

and J is the $n \times n$ matrix with ones on the antidiagonal.

Theorem 5.6. *For each n and prime power q, there is a generalized DFT for $G = Sp_{2n}(\mathbb{F}_q)$ that uses $O(|G|^{\omega/2+\epsilon})$ operations for all $\epsilon > 0$.*

Proof. Let L, U, D be the lower-triangular (with ones on the diagonal), upper-triangular (with ones on the diagonal), and diagonal subgroups of $\mathrm{GL}_{2n}(\mathbb{F}_q)$, respectively. We view our group G as a subgroup of $\mathrm{GL}_{2n}(\mathbb{F}_q)$ as well. It is well known that the order of G is

$$q^{n^2} \prod_{i=1}^{n} (q^{2i} - 1) \leq q^{2n^2+n}.$$

Now apply Theorem 5.1 with $H_1 = L \cap G, H_2 = D \cap G$, and $H_3 = U \cap G$. We note that H_1 and H_3 are p-groups and H_2 is Abelian (as before). Also, H_1H_2 is a subgroup, and $H_1 \cap H_2 = \{1\}$ and $H_1H_2 \cap H_3 = \{1\}$.

It remains to bound the sizes of H_1, H_2, H_3. To lower bound the size of H_3, consider the following subgroups of $\mathrm{GL}_{2n}(\mathbb{F}_q)$,

$$\begin{aligned} H &= \left\{ \left(\begin{array}{c|c} I_n & M \\ \hline 0 & I_n \end{array} \right) : M \in \mathbb{F}_q^{n \times n} \right\} \\ K &= \left\{ \left(\begin{array}{c|c} A & 0 \\ \hline 0 & B \end{array} \right) : A, B \text{ upper triangular } n \times n \text{ matrices with ones on the diagonal} \right\}. \end{aligned}$$

One can verify that $H \cap G$ is the subgroup in which M is a persymmetric matrix (symmetric about the anti-diagonal), and thus this subgroup has order $q^{n(n+1)/2}$. Similarly, one can verify that $K \cap G$ is the subgroup in which A is an arbitrary upper-triangular matrix with ones on the diagonal and $B = J(A^T)^{-1}J$. Thus this subgroup has order $q^{n(n-1)/2}$. We have

$$(H \cap G)(K \cap G) \subseteq H_3$$

and so $|H_3| \geq q^{n(n+1)/2+n(n-1)/2} = q^{n^2}$. A symmetric argument shows that $|H_1|$ has the same order. It is also easy to verify that $|H_2| = (q-1)^n$. In the notation of Theorem 5.1, we have

$$r = O\left(\frac{|G| \log |G|}{|H_1||H_2||H_3|} \right) \leq O\left(\frac{q}{q-1} \right)^n ((n^2+n) \log q),$$

which can be absorbed into the $|G|^\epsilon$ term. □

5.3 The Orthogonal Groups $O_n(\mathbb{F}_q)$

An orthogonal group of dimension n over $\mathbb{F}_q$ is a subgroup of invertible matrices that preserve a nondegenerate symmetric quadratic form. There are several inequivalent quadratic forms and thus several non-isomorphic orthogonal groups. For simplicity, we work out only one case (the "plus type" orthogonal group of even dimension, in odd characteristic). A similar analysis can be easily carried out for the other non-isomorphic orthogonal groups. In our case, concretely, we may take $O_n(\mathbb{F}_q)$ to be the set of all matrices $A \in \mathrm{GL}_n(\mathbb{F}_q)$ such that

$$A^T Q A = Q, \text{ where } Q = \begin{pmatrix} 0 & J \\ J & 0 \end{pmatrix},$$

and J is the $n/2 \times n/2$ matrix with ones on the antidiagonal.

Theorem 5.7. *For each even n and odd prime power q, there is a generalized DFT for $G = O_n(\mathbb{F}_q)$ specified via the above quadratic form that uses $O(|G|^{\omega/2+\epsilon})$ operations for all $\epsilon > 0$.*

Proof. Let L, U, D be the lower-triangular (with ones on the diagonal), upper-triangular (with ones on the diagonal), and diagonal subgroups of $\mathrm{GL}_n(\mathbb{F}_q)$, respectively. We view our group G as a subgroup of $\mathrm{GL}_n(\mathbb{F}_q)$ as well. It is well known that the order of G is at most $2q^{(n^2-n)/2}$.

Now apply Theorem 5.1 with $H_1 = L \cap G, H_2 = D \cap G$, and $H_3 = U \cap G$. We note that H_1 and H_3 are p-groups and H_2 is Abelian (as before). Also, H_1H_2 is a subgroup, and $H_1 \cap H_2 = \{1\}$ and $H_1H_2 \cap H_3 = \{1\}$.

It remains to bound the sizes of H_1, H_2, H_3. To lower bound the size of H_3, first consider the following subgroups of $\mathrm{GL}_n(\mathbb{F}_q)$,

$$H = \left\{ \left(\begin{array}{c|c} I_{n/2} & M \\ \hline 0 & I_{n/2} \end{array} \right) : M \in \mathbb{F}_q^{n/2 \times n/2} \right\}$$

$$K = \left\{ \left(\begin{array}{c|c} A & 0 \\ \hline 0 & B \end{array} \right) : A, B \text{ upper tri. with ones on the diagonal} \right\}.$$

One can verify that $H \cap G$ is the subgroup in which M is a "skew-persymmetric" matrix (skew-symmetric about the anti-diagonal), and thus this subgroup has order $q^{((n/2)^2-(n/2))/2}$. Similarly, one can verify that $K \cap G$ is the subgroup in which A is an arbitrary upper-triangular matrix with ones on the diagonal and $B = J(A^T)^{-1}J$. Thus this subgroup has order $q^{((n^2)^2-(n/2))/2}$. We have

$$(H \cap G)(K \cap G) \subseteq H_3,$$

and so $|H_3| \geq q^{(n/2)^2-(n/2)}$. A symmetric argument shows that $|H_1|$ has the same order. It is also easy to verify that $|H_2| = (q-1)^{n/2}$. In the notation of Theorem 5.1, we have

$$r = O\left(\frac{|G| \log |G|}{|H_1||H_2||H_3|}\right) \leq O\left(\frac{q}{q-1}\right)^{n/2} ((n^2 - n) \log q/2),$$

which can be absorbed into the $|G|^\epsilon$ term. □

We note that in all of the cases just considered in Sections 5.1, 5.2, 5.3, one obtains the same results for the special or projective (or both) variants by following essentially the same argument. To obtain results for the projective cases, we observe that quotient-ing all of the groups in our decomposition by the center can only change the operation count by a factor of some constant multiple of the size of the center, which in these cases is itself a constant.

Finally, we note that Theorem 5.2 and the surrounding discussion imply

Theorem 5.8. *Let G be a finite simple group. Then there is a fast DFT over G that uses $O(|G|^{\omega/2+\epsilon})$ operations for all $\epsilon > 0$.*

Proof. As noted in the discussion before and after Theorem 5.2, all finite simple groups of Lie type (except the Tits group) have a split (B, N)-pair, and Weyl group of order $|G|^{o(1)}$, so Theorem 5.2 yields exponent $\omega/2$ algorithms for these families. By the Classification Theorem, the only other infinite families of finite simple groups are the alternating group and the Abelian groups, both of which have exponent one algorithms. The sporadic groups and the Tits group are a finite set of exceptions that can be handled by choosing the constant in the big-oh notation sufficiently large. □

6 A NEW EXPONENT UPPER BOUND FOR ALL FINITE GROUPS

In this section, we prove a structural result for all finite groups that allows us to make use of the reduction in Theorem 3.7. Just as Lev's theorem regarding a large single subgroup allows one to use the single subgroup reduction of Section 2 to obtain a non-trival upper bound for all finite groups, the following theorem gives a *pair* of subgroups for use in the reduction of Theorem 3.7.

Theorem 6.1. *There exists a monotone increasing function $f(x) \leq 2^{c\sqrt{\log x} \log\log x}$ for a universal constant $c \geq 1$, for which the following holds: Every finite group G that is not a p-group has proper subgroups H, K satisfying $|HK| \geq |G|/f(|G|)$.*

Proof. If G is simple, then by the Classification Theorem, we have several cases:

- G is cyclic of prime order. This case cannot arise since G is not a p-group.
- G is the alternating group A_n. Then we choose $H = A_{n-1}$ and $K = \{1\}$, and we have $|HK| \geq |G|/n$, so as long as $f(x) > \log x$, the theorem holds.
- G is a finite group of Lie Type. Then G has a (B, N) pair (the Tits Group is an exception; it does not have a (B, N) pair, but it is a single finite group so it can be treated along with the sporadic groups in the next case). Let $W = N/(B \cap N)$ be the Weyl group, and from the axioms of a (B, N) pair, we have that the double cosets $B\overline{w}B$ with $w \in W$ cover G (the $\overline{w}$ denotes a lift to $N \subseteq G$). Thus there is some double coset $B\overline{w}B$ of size at least $|G|/|W|$. Taking H to be the conjugate subgroup $\overline{w}B\overline{w}^{-1}$ and $K = B$, we see that $|HK| = |B\overline{w}B| \geq |G|/|W|$. Now we verify that we can choose f as specified in the theorem statement, so that for each of the families in Figure 2, $f(|G|) > |W|$.
- G is one of the sporadic groups. Let C be the largest order of a sporadic group. Then by choosing $f(x) > C$, the theorem holds for $H = K = \{1\}$ in this case.

If G is not simple, then let N be a maximal normal subgroup of G, so that G/N is simple. We have two cases:

- G/N is a p-group. Since G is not a p-group, we have that $|G| = mp^k$ for $m > 1$ and $(m, p) = 1$. Let P be a p-Sylow subgroup of G. Then $|P| = p^k$, and $|N| = mp^{k'}$ for some $k' < k$. Then $NP = G$ and both N and P are proper subgroups.
- G/N is a simple group that is not a p-group. Then apply the previous case analysis for simple groups to obtain $H/N, K/N$, proper subgroups of G/N for which $|(H/N)(K/N)| \geq |G/N|/f(|G/N|)$. But then H, K are proper subgroups of G and

$$|HK| = |(H/N)(K/N)||N| \geq |G/N||N|/f(|G/N|) = |G|/f(|G/N|) \geq |G|/f(|G|),$$

where the last inequality used the monotonicity of f. □

Now we can use this theorem in a recursive algorithm that switches between the single subgroup reduction and the double subgroup reduction, as follows:

Theorem 6.2. *For every finite group G, there is an exponent $\frac{\omega-2+\sqrt{\omega^2-4\omega+36}}{4}$ algorithm computing the DFT with respect to G when $\omega \leq \frac{1+\sqrt{17}}{2}$, or exponent $\frac{4\omega}{4+\omega}$ when $\omega \geq \frac{1+\sqrt{17}}{2}$. In particular, when $\omega = 2$, the exponent is $\sqrt{2}$.*

To visualize these bounds, refer to Figure 3.

Proof. We describe our general strategy before formally analyzing the complexity. For each possible value of ω, we pick a threshold β as a function of ω. This threshold will be used to switch between the single subgroup and the double subgroup reductions.

Fix G. Consider the following recursive algorithm. If G is a p-group, then we apply Theorem 4.2. If G is the trivial group, then the DFT is trivial as well. Otherwise, let H, K be the subgroups guaranteed by Theorem 6.1. If $|H|, |K|$ are both at most $|G|^\beta$, then we apply Theorem 3.7 (the double subgroup reduction). Otherwise one of H, K has size at least $|G|^\beta$ (without loss of generality, assume it's H) and we apply Corollary 2.2 (the single subgroup reduction).

Let us now analyze the operation count in terms of β. After this analysis, we will pick the optimal β as a function of ω to minimize the operation count.

For this purpose, set $\delta = \min\{\epsilon, 0.1, \frac{0.1\epsilon}{\beta}\}$, and give names to some constants:

- Let A_δ be the constant hidden in the $O(|G|^{\omega/2+\delta} + (|H||K|)^{\omega/2+\delta})$ notation of Theorem 3.7.
- Let B_δ be the constant hidden in the $[G : H]^2 \cdot O(|H|^{\omega/2+\delta})$ notation of Corollary 2.2.
- Let B be the constant hidden in the $O(|G| \log |G|)$ notation of Theorem 4.2.

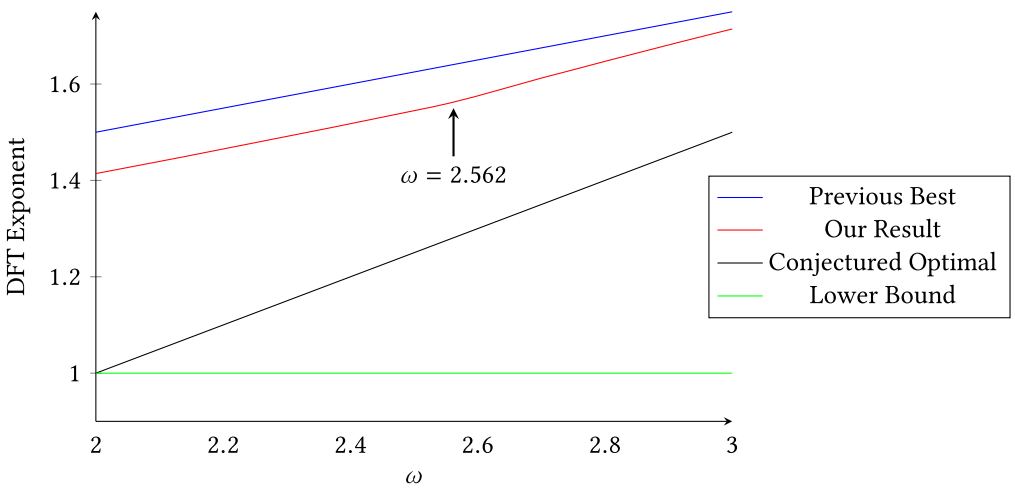

Fig. 3. Upper bound in Theorem 6.2 as a function of ω. The previous best bound is from Theorem 2.3. Assuming that some dependence on fast matrix multiplication is necessary, $w/2$ is a reasonable conjecture for the optimal dependence. Exponent one is of course a trivial lower bound.

Let $T(n)$ denote an upper bound on the running time of this recursive algorithm for any group G of order n. For each fixed $\epsilon > 0$, we will prove by induction on n that, for a universal constant C_ϵ,

$$T(n) \leq C_\epsilon n^{\alpha+\epsilon} \log^2 n, \tag{6}$$

where α is determined by β and ω. This clearly holds for the base case of a p-group or the trivial group, provided $C_\epsilon > B$ and $\alpha \geq 1$.

By selecting a sufficiently large universal constant C_ϵ, we may assume that $|G|$ is at least some fixed constant size, say, $C_\epsilon^{1/2}$ (since for smaller G the trivial algorithm will fall within the claimed time bound of Equation (6)). Hence, we may assume that $2^{c\sqrt{\log|G|}\log\log|G|} \cdot O(\log|G|)$ term in the notation of Theorem 3.7 is bounded above by $|G|^{\epsilon/10}$.

In the case where we apply Theorem 3.7, the cost is at most

$$\left(|H| \cdot T(|K|) + |K| \cdot T(|H|) + A_\delta(|H||K|)^{\omega/2+\delta}\right) \cdot |G|^{\epsilon/10},$$

where $|H|, |K| \leq |G|^\beta$. Applying the induction hypothesis, we obtain

$$\begin{aligned} T(n) &\leq 2C_\epsilon \left(n^\beta n^{\beta(\alpha+\epsilon)} \log^2(n^\beta) + A_\delta n^{2\beta(\omega/2+\delta)}\right) \cdot n^{\epsilon/10} \\ &\leq (2C_\epsilon\beta^2 + A_\delta) \cdot n^{\max(\beta+\beta\alpha+\beta\epsilon, \omega\beta+2\beta\delta)+\frac{\epsilon}{10}} \log^2 n, \end{aligned}$$

which can be bounded above by $C_\epsilon n^{\alpha+\epsilon} \log^2 n$ as long as the following constraints are satisfied:

- $\beta < \frac{\sqrt{2}}{2} \approx 0.707$;
- $\alpha \geq \max(\frac{\beta}{1-\beta}, \omega\beta)$;
- $C_\epsilon > \frac{A_\delta}{1-2\beta^2}$.

In the case where we apply Corollary 2.2, the cost is at most

$$[G:H] \cdot T(|H|) + [G:H]^2 \cdot B_\delta |H|^{\omega/2+\delta},$$

where $|H| \geq |G|^{\beta}$ and hence $[G : H] \leq |G|^{1-\beta}$. If we set γ such that $|H| = |G|^{\gamma}$, and thus $\beta \leq \gamma \leq 1$, and apply the induction hypothesis, then we obtain

$$\begin{aligned} T(n) &\leq C_\epsilon n^{1-\gamma} n^{\gamma(\alpha+\epsilon)} \log^2(n/2) + B_\delta n^{2(1-\gamma)} n^{\gamma(\omega/2+\delta)} \\ &< C_\epsilon n^{\alpha+\epsilon} (\log n)(\log n - 1) + B_\delta n^{2-(2-\omega/2)\beta+\delta}, \end{aligned}$$

which is at most $C_\epsilon n^{\alpha+\epsilon} \log^2 n$ as long as the following constraints are satisfied:

- $\alpha \geq 2 - (2 - \omega/2)\beta$;
- $C_\epsilon \geq B_\delta$.

To recap, the above induction proof holds when

$$\alpha = \max\left(\frac{\beta}{1-\beta}, \omega\beta, 2 - \left(2 - \frac{\omega}{2}\right)\beta\right), \text{ and } \beta < \frac{\sqrt{2}}{2} \approx 0.707.$$

Now we solve for the β minimizing α, as a function of ω.

When $\omega \geq \frac{1+\sqrt{17}}{2} \approx 2.562$, the optimal is

$$\beta^* = \frac{4}{4+\omega}, \alpha^* = \frac{4\omega}{4+\omega}.$$

When $\omega \leq \frac{1+\sqrt{17}}{2} \approx 2.562$, the optimal is

$$\beta^* = \frac{10 - \omega - \sqrt{\omega^2 - 4\omega + 36}}{2(4-\omega)}, \alpha^* = \frac{\omega - 2 + \sqrt{\omega^2 - 4\omega + 36}}{4}.$$ □

7 CONCLUSIONS

There are two significant open problems that naturally follow from the results in this article. First, can one obtain exponent $\omega/2$ algorithms for all finite groups? This might be possible by proving a more sophisticated version of Theorem 6.1, which, for example, manages to upper bound $|H \cap K|$. Also of interest would be a proof of Theorem 6.1 that does not need the Classification Theorem.

A second question is whether the dependence on ω can be removed. Alternatively, can one show that a running time that depends on ω is necessary by showing that an exponent one DFT for a certain family of groups would imply $\omega = 2$?

REFERENCES

[1] Ulrich Baum. 1991. Existence and efficient construction of fast Fourier transforms on supersolvable groups. *Comput. Complex.* 1, 3 (1 Sep. 1991), 235–256.

[2] Thomas Beth. 1984. *Verfahren der schnellen Fourier-Transformation.* Teubner.

[3] P. Bürgisser, M. Clausen, and M. A. Shokrollahi. 1997. *Algebraic Complexity Theory.* Grundlehren der mathematischen Wissenschaften, Vol. 315. Springer-Verlag.

[4] Roger W. Carter. 1989. *Simple Groups of Lie Type.* Vol. 22. John Wiley & Sons.

[5] Michael Clausen. 1989. Fast generalized Fourier transforms. *Theor. Comput. Sci.* 67, 1 (1989), 55–63.

[6] Michael Clausen and Ulrich Baum. 1993. *Fast Fourier Transforms.* Wissenschaftsverlag.

[7] Michael Clausen and Paul Hühne. 2017. Linear time Fourier transforms of Sn-k-invariant functions on the symmetric group Sn. In *Proceedings of the 2017 ACM on International Symposium on Symbolic and Algebraic Computation (ISSAC'17).* ACM, New York, NY, 101–108. DOI: https://doi.org/10.1145/3087604.3087628

[8] Roger A. Horn and Charles R. Johnson. 1991. *Topics in Matrix Analysis.* Cambridge University Press.

[9] Chloe Ching-Yun Hsu and Chris Umans. 2018. A fast generalized DFT for finite groups of Lie type. In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'18)*, Artur Czumaj (Ed.). SIAM, 1047–1059. DOI: https://doi.org/10.1137/1.9781611975031.68

[10] John D. Lafferty and Daniel Rockmore. 1992. Fast Fourier analysis for SL_2 over a finite field and related numerical experiments. *Exper. Math.* 1, 2 (1992), 115–139. http://projecteuclid.org/euclid.em/1048709049

[11] François Le Gall. 2014. Powers of tensors and fast matrix multiplication. In *Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation*. ACM, 296–303.
[12] Arieh Lev. 1992. On large subgroups of finite groups. *J. Algebr.* 152, 2 (1992), 434–438.
[13] David Maslen and Daniel Rockmore. 1997. Separation of variables and the computation of Fourier transforms on finite groups, I. *J. Am. Math. Soc.* 10, 1 (1997), 169–214.
[14] David Maslen, Daniel N. Rockmore, and Sarah Wolff. 2016. The efficient computation of Fourier transforms on semisimple algebras. *Journal of Fourier Analysis and Applications* 24 (2016), 1377–1400.
[15] David Maslen, Daniel N. Rockmore, and Sarah Wolff. 2018. Separation of variables and the computation of Fourier transforms on finite groups, II. *J. Fourier Anal. Appl.* 24, 1 (2018), 226–284.
[16] David Keith Maslen. 1998. The efficient computation of Fourier transforms on the symmetric group. *Math. Comput.* 67, 223 (1998), 1121–1147. DOI: https://doi.org/10.1090/S0025-5718-98-00964-8
[17] David K. Maslen and Daniel N. Rockmore. 1997. Generalized FFTs – A survey of some recent results. In *Groups and Computation II*, Vol. 28. American Mathematical Society, 183–287.
[18] David K. Maslen and Daniel N. Rockmore. 2000. Double coset decompositions and computational harmonic analysis on groups. *J. Fourier Anal. Appl.* 6, 4 (2000), 349–388.
[19] Daniel Rockmore. 1997. Some applications of generalized FFTs. In *Proceedings of the 1995 DIMACS Workshop on Groups and Computation*. June, 329–369.
[20] Daniel N. Rockmore. 1995. Fast Fourier transforms for wreath products. *Appl. Comput. Harmon. Anal.* 2, 3 (1995), 279–292. DOI: https://doi.org/10.1006/acha.1995.1020
[21] Daniel N. Rockmore. 2002. Recent progress and applications in group FFTs. In *Proceedings of the Conference Record of the 36th Asilomar Conference on Signals, Systems and Computers, 2002*, Vol. 1. IEEE, 773–777.
[22] Wikipedia. 2017. List of Finite Simple Groups. Retrieved June 30, 2017 from https://en.wikipedia.org/w/index.php?title=List_of_finite_simple_groups&oldid=786516939.

Received April 2018; revised November 2018; accepted November 2018