Adapted Diameters and the Efficient Computation of Fourier Transforms on Finite Groups

David K. Maslen*
Max-Planck-Institut-fur-Mathematik
53225 Bonn
Germany

Daniel N. Rockmore†
Dept. of Math and Comp. Sci.
Dartmouth College
Hanover, NH 03755

Abstract

This paper introduces new techniques for the efficient computation of the Discrete Fourier Transform (DFT) for a finite group G and in so doing, relates the complexity of a finite group to its **adapted diameter**, relative to a given generating set and chain of subgroups. Consequently, we are able to show, for the first time, that the complexity of the DFT of a finite group is intimately related to group structure and thereby begin to link two major areas of research in computational group theory. In many particular cases, the resulting algorithms have potential applications for data analysis and signal processing. Given a chain of subgroups for a group G we introduce a technique which produces factorizations of group elements into short products of elements which commute with various subgroups along the chain. The commutativity properties of the factors is used to to show that for any irreducible matrix representation of the group these elements will have factorizations as highly structured sparse matrices. This allows a separation of variables style algorithm to be used in computing the associated DFT and consequent speedups follow immediately. In particular, this technique recovers the best known algorithms for the symmetric groups and wreath products and beyond that, dramatically improves on the known complexity of the DFT for the finite groups $GL(n,q)$ and gives first complexity results for all finite classical groups, finite groups of Lie type and groups with a (B,N)-pair.

1 Introduction

Recently, increased attention has begun to be paid to the problem of finding efficient algorithms for the computation of the Discrete Fourier Transform (DFT) for a finite group G. The abelian case has long been of interest and its "solution" in the form of the Cooley-Tukey Fast Fourier Transform (FFT) [14] and its many variants (cf. [19, 37, 36] and the many references contained therein) have been crucial to the development and utility of digital signal processing.

The initial motivation for developing FFT's for non-abelian groups was also driven by applications. To date, FFT's for nonabelian finite groups were found to be useful and necessary for new approaches to problems in data analysis [15], VLSI design [7], the design of matched filters [26] and efficient group convolution algorithms [10, 29]. In the continuous setting, applications to computer vision, geophysics and climate modeling have been identified and pursued [18, 23]. Conversely, as new algorithms are developed, new applications are sought. This symbiosis between application and theory continues.

Beyond their applicability, these algorithms are of intrinsic theoretical interest within the current effort towards the classification of finite groups according to upper bounds on the complexity of an associated DFT (cf. Section 2.1). Direct computation of any DFT for G requires at most $|G|^2$ operations. Whereas in the abelian case there is an essentially unique choice of DFT, for nonabelian groups there are an infinite number of possibilities and upper bounds on the complexity of the computation vary with choice of basis for the DFT. The complexity of the group is defined as the least upper bound over the complexities of all DFT's. It is conjectured that all finite groups have complexity $O(|G| \log^c |G|)$. To date, this has been shown to be true for many different classes of nonabelian groups ([11, 30, 31, 5]).

Until now, the development of FFT algorithms has focused almost exclusively on the representation theory of the various groups of interest, with little attention paid to the intrinsic structure of the group. The main tool in current use is the construction of **subgroup-**

*Partially supported as a Shapiro Visitor while at Dartmouth.
†This work supported in part by ARPA as administered by the AFOSR under contract DOD F4960-93-1-0567 as well as an NSF Math Sciences Postdoctoral Fellowship.

adapted bases which permit the DFT of the group G to be "reduced" to the computation of DFT's over a subgroup H, which are then "glued together" via multiplication by "twiddle factors" - which are precisely the evaluation of irreducible matrix representations at coset representatives for the subgroup. In the abelian case these twiddle factors are simply roots of unity and their multiplication does not affect the asymptotic complexity of the computation. However, in the nonabelian case, these twiddle factors become full matrices and without any information about their structure, require a complete matrix multiplication in order to be applied. The accumulation of these matrix multiplications severely degrades the efficiency of the algorithm. In particular cases, structure for these matrices has been discovered and consequent savings obtained, but to date, no general theory indicating how such structure might be uncovered has been presented. It is the purpose of this paper to exhibit such a theory and show its wide range of applicability.

In brief, the main idea of this paper is as follows. Given a chain of subgroups for a group G and subgroup-adapted bases for the irreducible representations of G, the goal is to factor the successive sets of coset representatives in terms of group elements which commute with various subgroups within the chain. Commutativity then implies that the irreducible matrix representations at these "commuting elements" are generally block diagonal matrices in which the blocks are themselves block matrices with scalar diagonal blocks. Consequently by applying the twiddle factors as a succession of matrices of these type, great savings are obtained. For many situations of interest (e.g. symmetric groups, finite classical groups, finite groups of Lie type) the groups are given by natural sets of generators and the resulting complexity of the DFT has an upper bound neatly encoded in terms of the **adapted diameter** of the group (cf. Section 2) for the given generating set and subgroup chain. The search for optimal DFT's then becomes a problem in minimizing various intrinsically group theoretic parameters.

By bringing the internal structure of the group to bear on the problem of group complexity, the results in this paper imply a closer connection than had previously been thought between the areas of FFT research and computational group theory. The latter subject, as pioneered by C. C. Sims, grew out of the necessity for the development of computational tools to aid the completion of the Classification of Finite Simple Groups. It has since then grown to encompass a huge wealth of both theoretical and practical algorithmic techniques for determining the structure of any finitely presented group. The volume [20] is a current source for many of the new developments in computational group theory and provides pointers to much of the important literature.

Section 2 presents the background and gives a brief explanation of the problem. In Section 3 we present the main result (Theorem 3.2 and its Corollary), deferring a discussion of its proof to Section 5, in order to proceed directly to some of the important consequences in Section 4. There we show how our techniques reobtain many of the best known FFT algorithms and beyond that are able to derive greatly improved complexity results for the matrix groups over finite fields $GL(n, q)$, as well as the first results for the other finite classical groups, finite groups of Lie type and more generally, finite groups with a (B, N)-pair. We close in Section 6 with a brief indication of possible improvements and generalizations.

Acknowledgement. Special thanks to Tom Hagedorn for patiently explaining his interesting recent work on multiplicities for restricted representations. Thanks also to Herr Prof. Michael Clausen for some very helpful conversations.

2 Preliminaries

2.1 DFT's and FFT's. The familiar "usual" or circular, or more generally, abelian **Discrete Fourier Transform** (DFT) and subsequent efficient reorganization via the Cooley-Tukey **Fast Fourier Transform** (FFT) [14] has a natural formulation in terms of the **representation theory** of cyclic or abelian groups. This larger framework is necessary for posing the general problem of efficient computation of DFT's for finite groups. What follows is a brief review of these ideas, including the notion of **subgroup-adapted** set of representations which is crucial for many of the constructions of FFT's for finite nonabelian groups. For a complete introduction to the subject Serre's book [32] is a good reference.

Recall that a (complex) **matrix representation** of a finite group G is a function ρ from G into $GL_d(\mathbf{C})$, the group of $d \times d$ invertible matrices with complex entries such that $\rho(st) = \rho(s)\rho(t)$ for every $s, t \in G$. In this case d is called the **degree** or **dimension** of the representation ρ, and is denoted d_ρ.

Two representations ρ_1 and ρ_2 are said to be **equivalent** if they differ only by a change of basis, so if there exists an invertible matrix A such that $\rho_1(s) = A^{-1}\rho_2(s)A$ for all $s \in G$. Notice that 1-dimensional matrix representations are uniquely determined by their equivalence class, while multidimensional representations have an infinite number of equivalent realizations.

A subspace $W \subset V = \mathbf{C}^d$ is said to be G-**invariant** if for all $s \in G$, $\rho(s)W \subset W$. The representation ρ is said to be **irreducible** if $V = \mathbf{C}^d$ has no G-invariant subspaces other than the trivial subspaces $\{0\}$ and V and **reducible** otherwise. Up to equivalence there are only a finite number of irreducible representations of any finite group - in fact there are as many as there are conjugacy classes in the group. Irreducible representations are the fundamental building blocks of all representations of a finite group. That is to say that any representation is equivalent to the direct sum of irreducible representations, where the direct sum of two representations is the matrix direct sum of the representations.

There are several equivalent definitions of the discrete Fourier transform for a finite group [10, 7, 16]. The following is the most convenient for this paper.

Definition 2.1. (Discrete FourierTransform) *Let G be a finite group, and f be a complex valued function on G.*

(a) For a matrix representation ρ G, the **Fourier transform of** f **at** ρ, *denoted $\hat{f}(\rho)$ is the matrix sum,*

$$\hat{f}(\rho) = \sum_{s \in G} f(s)\rho(s).$$

(b) Assume $\mathcal{R} = \{\rho_1, \ldots, \rho_k\}$ is a complete set of irreducible matrix representations of G. The **Discrete Fourier transform of** f *(with respect to $\mathcal{R}$), denoted $DFT(f)$, is the set of Fourier transforms of f at the representations in $\mathcal{R}$,*

$$DFT(f) = \{\hat{f}(\rho_1), \ldots, \hat{f}(\rho_k)\}.$$

A DFT of f determines f through the Fourier inversion formula,

$$f(s) = \frac{1}{|G|} \sum_{\rho \in \mathcal{R}} \operatorname{trace}\Big(\hat{f}(\rho)\rho(s^{-1})\Big). \qquad (2.1)$$

Example: The "usual" DFT. The irreducible matrix representations of the cyclic group $\mathbf{Z}/n\mathbf{Z}$, are all one dimensional. For each integral j with $0 \le j \le n-1$, define the representation, ζ_j, by $\zeta_j(k) = \exp(\frac{2\pi i j k}{n})$ where k is in $\mathbf{Z}/n\mathbf{Z}$. The set of such representations is a complete set of inequivalent irreducible representations for $\mathbf{Z}/n\mathbf{Z}$ and the corresponding DFT is the usual discrete Fourier transform.

Definition 1. (Complexity) *Assume G is a finite group, and $\mathcal{R}$ is any set of matrix representations of G. Let $T_G(\mathcal{R})$ denote the minimum number of operations needed to compute $DFT(f, \mathcal{R})$ via a straight line program for an arbitrary complex function on G. $T_G(\mathcal{R})$ is called the* **complexity** *of the DFT for the set $\mathcal{R}$. Furthermore, define the* **complexity of the group** *G to be*

$$\mathcal{C}(G) = min\{T_G(\mathcal{R})\}$$

where $\mathcal{R}$ varies over complete sets of inequivalent irreducible matrix representations of G.

The computational model used here is a common one in which an operation is defined as a single complex multiplication followed by a complex addition.

Elementary representation theory shows that the sum of the squares of the degrees of a complete set of irreducible representations of G is equal to $|G|^2$ [32]. Consequently direct computation of any DFT shows

$$|G| \le T_G \le |G|^2.$$

Remark. Another common interpretation of the DFT is as a change of basis for $L(G)$, from the basis of point masses on G, to a basis of matrix coefficients making up a complete set of inequivalent irreducible representations. When this approach is adopted, the complexity of the DFT can be measured as the c-**linear complexity** of the DFT matrix [6]. The c-linear complexity of a group G, is defined to be the minimum c-linear complexity of any DFT matrix for G. Assuming a choice of unitary representations, the results stated here can all be translated into statements about the 2-linear complexity of finite groups.

Fast Fourier transforms (FFT's) are algorithms for computing DFT's efficiently. As remarked earlier, there are an infinite number of matrix representations equivalent to any given multidimensional matrix representation; these correspond to changing bases in $\mathbf{C}^d$. The complexity of the discrete Fourier transform may vary with the representation even amongst equivalent representations. Subgroup-adapted sets of representations permit the computation of a DFT on a group G to be reduced to the computation of a DFT on a chosen subgroup H. The idea is quite simple to explain.

If H is a subgroup of G and $Y \subset G$ is a set of coset representatives for G/H (so G can be factored as the disjoint union of subsets $yH = \{yh \ : \ h \in H\}$ for all $y \in Y$) then for any representation ρ of G, we can use the relation $\rho(ab) = \rho(a)\rho(b)$ to obtain a factorization of $\hat{f}(\rho)$ by

$$\begin{aligned} \hat{f}(\rho) &= \textstyle\sum_{s \in G} f(s)\rho(s) \\ &= \textstyle\sum_{y \in Y} \rho(y) \sum_{t \in H} f_y(t)\rho(t) \end{aligned} \qquad (2.2)$$

where for each $y \in Y$, f_y is the function on H defined by $f_y(t) = f(yt)$ for all $t \in H$. Consequently, with this notation we can rewrite $\hat{f}(\rho)$ as a sum of DFT's on H,

$$\hat{f}(\rho) = \sum_{y \in Y} \rho(y) \widehat{f_y}(\rho \downarrow H)$$

where $\rho \downarrow H$ is the representation on H given by **restricting** ρ to elements of the subgroup H.

Thus, if the DFT's $\widehat{f_y}(\rho \downarrow H)$ were computed for a complete set of irreducible representations ρ of G, then these matrices could then be glued together by the "twiddle factors"[1] $\rho(y)$ to build all the $\hat{f}(\rho)$ and the DFT of f.

Subgroup adapted representations permit the restricted transforms $\hat{f}_y(\rho \downarrow H)$ to be computed quickly. In general, a restricted representation, $\rho \downarrow H$ will be reducible, even when ρ is irreducible. Consequently, $\rho \downarrow H$ will be *equivalent* to the direct sum of irreducible representations of H, although not necessarily *equal*. **H-adapted representations** guarantee that (1) the restriction of any representation of G to H is equal to a direct sum of irreducible representations of H and furthermore (2) that equivalent irreducible blocks among the restricted representations are in fact equal. To briefly illustrate, suppose that $\rho_1, \ldots, \rho_k$ are a complete set of irreducible representations for G. Then this set will be H-adapted if for every $t \in H$,

$$\rho_i(t) = \begin{pmatrix} B_{i,1}(t) & 0 & \cdots & 0 \\ 0 & B_{i,2}(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_{i,r_i}(t) \end{pmatrix}$$

where the $B_{i,j}(t)$ are irreducible representations of H such that if $B_{i,j}$ and $B_{l,m}$ give rise to equivalent irreducible representations then in fact, $B_{i,j}(t) = B_{l,m}(t)$.

Consequently, it is easy to see that an algorithm for computing a DFT of any function f on G with respect to an H-adapted basis for the irreducible representations of G is to (1) choose a set of coset representatives Y for G/H, compute the the DFT's on H for each f_y, and (2) build the restricted transforms $\widehat{f_y}(\rho \downarrow H)$ - these will be block diagonal with the individual Fourier transforms of f_y making up the blocks and then (3) compute the products $\rho(y)\widehat{f_y}(\rho \downarrow H)$ and add them together. Thus, the following theorem is obtained.

Theorem 2.1. (Diaconis and Rockmore) *Let H be a subgroup of G and let $Y \subset G$ be a set of coset representatives for G/H. Then the complexity of the DFT on G at a complete H-adapted set of inequivalent irreducible representations of G is related to the corresponding DFT on H by*

$$T_G \leq \frac{|G|}{|H|} T_H + M_G(Y) \tag{2.3}$$

where

$$M_G(Y) \;=\; \text{The number of operations needed to compute } \textstyle\sum_{y \in Y} \rho(y) F_y(\rho) \text{ for arbitrary } d_\rho \times d_\rho \text{ matrices } F_y(\rho). \tag{2.4}$$

The inequality (2.3) can be viewed as recurrence, bounding the complexity of a group in terms of a subgroup. This generalizes immediately to a chain of subgroups for G,

$$K_n > K_{n-1} > \ldots > K_0 \tag{2.5}$$

We say that a complete set of irreducible matrix representations is adapted to the chain provided that it is K_i-adapted for each subgroup, K_i, in the chain. It is easy to see that this in turn implies that for each K_i, the set of irreducible representations occurring in the restrictions to K_i is then K_j-adapted for each $j \leq i$. Theorem 2.1 now generalizes immediately.

Theorem 2.2. *Let G have the chain of subgroups (2.5) with $G = K_n$ and suppose that for $i = 1, \ldots, n$, Y_i is a set of coset representatives for K_i/K_{i-1}. Then for a set of matrix representations adapted to this chain we have*

$$T_G \leq |G| \left(\frac{1}{|K_0|} T_{K_0} + \sum_{i=1}^{n} \frac{1}{|K_i|} M_{K_i}(Y_i) \right). \tag{2.6}$$

Consequently, one approach to minimizing an upper bound of T_G is to find an efficient way of evaluating sums of the form $\sum_{y \in Y} \rho(y) A_y$.

There are several choices available to perform this minimization. The subgroup chain can be varied, as can the choice of coset representatives in order to obtain matrices $\rho(y)$ with useful computational properties. Another idea is to attempt to use the properties of the matrix elements of $\rho(y)$ as special functions on the set Y. The former idea is akin to the method of separation of variables and uses more the large scale internal structure of the group. As we will soon show, in many cases of interest this can be quantified in terms of relations between subgroup chains and generating sets.

Remark. The requirement of adaptability does not limit us. For any finite group G generated by a subset S, a complete set of irreducible representations can always be constructed in polynomial time [2]. Using techniques for decomposing representations into their irreducible constituents [3], these representations can then be transformed so that they are subgroup-adapted.

[1] The terminology "twiddle factor" comes from the usual signal processing situation in which G is an abelian group. Then all irreducible representations are one-dimensional and the matrices $\rho(y)$ are then simply roots of unity.

2.2 Adapted diameters, subgroup chains and generators. A common theme in the algorithms presented here is that of a factorization of coset representatives as products of elements which have "nice" commutativity properties relative to a given subgroup chain. Alternatively, it is perhaps useful to focus on the elements instead, necessarily generators for the group, and derive some general upper bounds from this perspective. In particular, this relates to the notion of the diameter of the group relative to this generating set.

Assume G is a finite group, and let $S \subset G$, be a set of **generators** for G. Recall that the **diameter** of G with respect to S is the maximum length of a product of elements in S required to express any element of G, i.e.

$$\gamma_S(G) = \min\Big\{n \geq 0 : \bigcup_{0 \leq i \leq k} S^i = G\Big\}.$$

Let a chain of subgroups of G be given,

$$K_m > \cdots > K_0. \tag{2.7}$$

Then S is said to be a **generating set for the chain of subgroups** if $S \cap K_j$ generates a set of coset representatives for K_j/K_0 for each j. When the subgroup chain contains both the whole group, G, and the trivial subgroup, 1, a generating set for the chain (2.7) is called a **strong generating set** for G with respect to the chain of subgroups (2.7). Strong generating sets arise naturally in the context of many algorithmic issues in computational group theory [33]. In particular, fast algorithms for their construction for stabilizer subgroup chains in permutation groups are a cornerstone for many important techniques [1].

The notion of diameter has a natural extension to subgroup chains and strong generating sets. Given a generating set S for a chain of subgroups of G, consider the sequence of integers,

$$\gamma_j = \min\Big\{l \geq 0 : \bigcup_{i \leq l} (S \cap K_j)^i \cdot K_{j-1} = K_j\Big\}. \tag{2.8}$$

Thus γ_j is the maximum length of a product of elements in $S \cap K_j$ needed to construct the coset representatives of K_j/K_{j-1}. Now define the **adapted diameter** of the chain $\{K_i\}$ relative to S to be

$$\gamma = \gamma(\{K_i\}, S) = \gamma_1 + \cdots + \gamma_n. \tag{2.9}$$

Notice that in this notation $\gamma_j = \gamma(K_j > K_{j-1}, S)$ and so is the adapted diameter of the chain $K_j > K_{j-1}$. When the chain of subgroups contains both G and $\{1\}$ we call (2.9) the **adapted diameter of** G, relative to S and $\{K_i\}$.

The adapted diameter of G reflects the way in which the Cayley graph on S can be "grown" through a chain of subgroups.

3 The Improved FFT

Suppose, as in Section 2.2, that G is generated by a subset S and furthermore, that a chain of subgroups of G is given

$$G = K_n > K_{n-1} > \cdots > K_0. \tag{3.10}$$

As remarked in the discussion leading to Theorem 2.1, in order to compute a DFT for a complete set of irreducible representations of G adapted to (3.10), we need an efficient way to evaluate the matrix products $\rho(y)\widehat{f_y}(\rho)$ as y varies over a complete set of coset representatives of G/H and ρ varies over a complete set of irreducible representations of G. The idea behind effecting this is the crucial observation that $\rho(x)$ for any element $x \in G$ which commutes with any subgroup in the chain will be a highly-structured and sparse matrix. Thus, if coset representatives y can be factored in terms of such elements the application of $\rho(y)$ to any $d_\rho \times d_\rho$ matrix can be achieved as a sequence of sparse matrix multiplications, thereby inducing great savings. We proceed by stating these results and postpone sketches of the proofs until Section 5.

Given any g in G, let $i^+(g)$ be such that $K_{i^+(g)}$ is the smallest subgroup in the chain containing g, and similarly, let $i^-(g)$ be such that $K_{i^-(g)} \leq K_{i^+(g)}$ is the largest subgroup of $K_{i^+(g)}$ in the chain (3.10) which commutes with g. Let $\mathcal{M}(g)$ be the maximum multiplicity (cf. Section 2.1) of any irreducible representation of $K_{i^-(g)}$ in the restriction of an irreducible representation of $K_{i^+(g)}$, and for a subset of G let $\mathcal{M}(S)$ denote the maximum of $\mathcal{M}(g)$ on S.

Lemma 3.1. *Assume ρ is an chain-adapted irreducible matrix representation of G, and F is any $d_\rho \times d_\rho$ matrix. Then for any $s \in G$, the matrix multiplication $\rho(x) \cdot F$ may be performed in less than than $\mathcal{M}(y)d_\rho^2$ operations.*

Using Lemma 3.1 it is straightforward to bound the quantity $M_G(Y)$ appearing in Theorem 2.1. In particular, suppose that $y \in Y$ has a factorization $y = s_1 \cdots s_k$ with $s_i \in S$. Then using the fact that $\sum d_\rho^2 = |G|$, we obtain

$$M_G(y) \leq k \cdot \mathcal{M}(S) \cdot |G|.$$

Since any coset of G/H has a representative which is a product of at most $\gamma(G > H, S)$ elements of S (cf. Section 2.2), by taking Y to be a set of coset representatives of G/H of minimal lengths in the elements of S yields

$$M_G(Y) \leq \mathcal{M}(S)\gamma(S, G > H)\,|G/H| \cdot |G|.$$

This is summarized in the following Theorem and its Corollary.

THEOREM 3.1. *Let G be a finite group with subgroup H and let $S \subset G$ generate a set of coset representatives for G/H. Assume that $G = K_m > \ldots > K_0$ is a chain of subgroups of G such that $H = K_i$ for some i. Then relative to a complete set of irreducible representations of G which are chain-adapted, we have*

$$(3.11)\quad \frac{\mathcal{C}(G)}{|G|} \leq \frac{T_G}{|G|} \leq \frac{|G|}{|H|}\gamma(S, G > H)\mathcal{M}(S) + \frac{T_H}{|H|}$$

COROLLARY 3.1. *Assume G is a finite group, $S \subset G$ is a strong generating set for G relative to the chain of subgroups $G = K_n > K_{n-1} > \cdots > K_0 = 1$. Then the number of operations needed to compute the Fourier transform of any given complex function on G at any chain-adapted set of representations satisfies*

$$\begin{aligned} T_G &\leq |G| \textstyle\sum_{i=1}^{m} \frac{|K_i|}{|K_{i-1}|}\gamma(S, K_i > K_{i-1})\mathcal{M}(S \cap K_i) \\ &\leq |G|\, \kappa . \gamma(S, \{K_i\}_{i=0}^{n})\mathcal{M}(S) \end{aligned}$$

where κ is the maximum of $|K_i/K_{i-1}|$ for $1 \leq i \leq n$. Consequently, the complexity of G, $\mathcal{C}(G)$ is similarly bounded.

4 Applications

We postpone the proof of Lemma 3.1 in order to move directly to some of its important applications. We first show how our general machinery reobtains the best known FFT's for some abelian groups, the symmetric groups and their wreath products and then move on to derive new results for the general linear groups over finite fields as well as their various generalizations.

4.1 Finite abelian groups. Corollary 3.1 immediately gives us some well known results bounding the complexity of the Fourier transform on finite abelian groups.

Assume A is a finite abelian group. Then the irreducible representations of A are all one dimensional so any choice of bases for a complete set of irreducible representations is adapted with respect to any chain of subgroups of G. We shall take $S = A$ to be our generating set; it is immediate that $\mathcal{M}(S) = 1$. Let $A = K_n > \ldots K_0 = 1$ be any chain of subgroups of A. Then the adapted diameter of this chain with respect to S is simply n. In particular, the adapted diameter of the two subgroup chain $K_i > K_{i-1}$ is $\gamma(A, K_i > K_{i-1}) = 1$. Applying Corollary 3.1 yields

$$(4.12)\qquad \frac{1}{|A|}T_A \leq \sum_{i=1}^{n} \frac{|K_i|}{|K_{i-1}|}$$

The left-hand side of (4.12) is a sum of factors of $|A|$ whose product is equal to $|A|$, and it is easy to see that such a sum is minimized precisely when each term $|K_i| \,/\, |K_{i-1}|$ is prime. One can always find such a chain in an abelian group; any chain of subgroups may be refined to such a chain. Thus we obtain

THEOREM 4.1. *Assume A is a finite abelian group whose order has the prime factorization $|A| = p_1^{r_1} \ldots p_m^{r_m}$. Then complexity of the Fourier transform on A satisfies*

$$T_A \leq |A| \sum_{i=1}^{m} r_i p_i .$$

This is essentially the well-known Cooley-Tukey FFT [14].

4.2 FFT's for S_n and its wreath products. For the symmetric group S_n consider the subgroup chain

$$S_n > S_{n-1} > \cdots > S_1 = \{1\}$$

where S_k is identified with the subgroup of S_n of elements fixing the points $k+1, \ldots, n$. Take as generating set the **pairwise-adjacent transpositions**

$$S = \{t_2, \ldots t_n\}$$

where t_j denotes the transposition $(j-1, j)$. Then note that $t_j \in S_j$ and commutes with S_k for $k < j-1$. Thus, in the notation of Section 3.1 $i^+(t_j) = j$ and $i^-(t_j) = j-2$. Furthermore, it is an easily derived fact from the combinatorics of Young tableaux, (cf. [24]) that the maximum multiplicity occurring in the restriction of any irreducible representation from S_j to S_{j-2} is 2, so that $\mathcal{M}(t_j) = 2$.

Lastly, note that coset representatives for S_k/S_{k-1} are given by the elements

$$1, t_k, t_{k-1}t_k, \ldots, t_2 \cdots t_k .$$

Thus $\gamma(S, S_k > S_{k-1}) = k$. Plugging this data into Theorem 3.1 gives

THEOREM 4.2.

$$\mathcal{C}(S_n) \leq T_{S_n} \leq \frac{2}{3}n(n+1)^2 n!$$

This is of order $n!(\log n!)^3$. Note that in this case the bases given by either **Young's orthogonal form** or **Young's seminormal form** are adapted for the chain of subgroups for S_n. The resulting algorithm is precisely the best known for computing the DFT for S_n [13].

For wreath products of the form $G[S_n]$, a similar construction works. Wreath products are of interest in data analysis as the symmetry groups of nested designs and in structural chemistry as the automorphism groups of non-rigid molecules. They are often studied as the automorphism groups of graphs obtained by "composition" (cf. [22]).

Elements of this group may be described by pairs $(f;\sigma)$ where $f : \{1,\ldots,n\} \longrightarrow G$, so $f(i)$ is the automorphism group acting on the i^{th} subgraph and $\pi \in S_n$ with a multiplication defined by $(f;\pi)\cdot(g;\sigma) = (f\cdot g^{\pi};\pi\sigma)$ where $f\cdot g^{\pi}(\mathrm{j}) = f(j)g^{\pi}(j)$ and g^{π} is defined by $g^{\pi}(j) = g(\pi^{-1}(j))$. In this notation it is clear that S_n sits naturally as a subgroup of $G[S_n]$ as is the product G^n - which is in fact a normal subgroup. A thorough but accessible treatment of wreath products may be found in [27].

Then a natural chain of subgroups for $G[S_n]$ is

$$(4.13) \quad G[S_n] > G\times G[S_{n-1}] > G[S_{n-1}] > \cdots$$

As before we let S denote the set of pairwise-adjacent transpositions in S_n, so S generates $G[S_n]$ modulo $G\times G[S_{n-1}]$ and the adapted diameter of the chain $G[S_n] > G\times G[S_{n-1}]$ relative to S is n. The transposition t_j lies in $G[S_j]$ and commutes with $G[S_{j-2}]$. It then can be shown that $\mathcal{M}(S)$ is $2(d_G)^2$, for d_G the maximum dimension of an irreducible representation of G. Hence we have

$$(4.14) \quad \begin{aligned} \frac{T_{G[S_n]}}{|G[S_n]|} &\le \frac{T_{G\times G[S_{n-1}]}}{|G|\,|G[S_{n-1}]|} + 2n^2(d_G)^2 \\ &\le \frac{T_{G[S_{n-1}]}}{|G[S_{n-1}]|} + \frac{T_G}{|G|} + 2n^2(d_G)^2 \end{aligned}$$

where the second inequality follows from the fact that in general, $T_{H\times K} \le |H|T_K + |K|T_H$ (cf. [10]). Applying the inequality (4.14) recursively gives us

THEOREM 4.3.

$$\mathcal{C}(G[S_n]) \le T_{G[S_n]} \le |G[S_n]|\left[\frac{2}{3}n(n+1)^2(d_G)^2 + n\frac{T_G}{|G|}\right]).$$

Given a subgroup chain for G one can construct a chain of subgroups of $G[S_n]$ refining the chain (4.13). Complete sets of irreducible matrix representations adapted to the subgroup chain (4.13) have been constructed and the above discussion recovers the best known algorithm for wreath products of the form $G[S_n]$ [31].

4.3 A new FFT for the general linear group over a finite field.

As usual, let $GL_n(q)$ denote the group of invertible $n\times n$ matrices with entries in the field of q elements where q is some prime power. For data analysis, these groups and their generalizations are of interest as the automorphism groups of the many designs based on finite geometries and codes.

Throughout this section all matrix groups are assumed to be over the finite field of q elements $\mathbf{F}_q$ so that $GL_n \equiv GL_n(q)$, etc.

THEOREM 4.4. *There is a positive constant, K, such that for any $n \ge 2$, $q \ge 2$, the DFT of a function defined on $GL_n(q)$ can be computed, using an adapted set of matrix representations for the chain of subgroups (4.15), in less than*

$$T_{GL_n} \le |GL_n(q)|\left[3n(2q)^{2n-1} + Kn(2q)^{2n-2}\right]$$

operations, and hence the complexity $\mathcal{C}(GL_n(q))$ is similarly bounded.

Proof. (*Sketch.*) To apply our ideas to these groups, we consider the chain of subgroups

$$(4.15)\ GL_n > P_{n-1} > GL_{n-1}\times GL_1 > GL_{n-1} > \cdots$$

where P_{n-1} is the subgroup of all block matrices of the form

$$(4.16) \qquad \left(\begin{array}{c|c} * & * \\ \hline 0\ \ldots\ 0 & b_n \end{array}\right)$$

for $b_n \in \mathbf{F}_q^{\times}$, and $GL_k \times GL_1$ is identified with the subgroup of block diagonal matrices $\mathrm{Diag}(A, b_k, I_{n-k+1})$ with A in GL_k and b_k in GL_1. [2]

In order to apply Theorem 3.1 we will describe a set of generators which have good commutativity properties relative to the subgroup chain (4.15).

For $i = 2,\ldots,n$ define the subgroups $A_i \cong GL_2$ consisting of block diagonal matrices with an arbitrary element of GL_2 in the $i-1, i$ block and all other diagonal elements equal to 1. We let our generating set, S, be the union of the A_i for $2 \le i \le n$.

Notice that A_i commutes with GL_{i-2}. Following the general philosophy then of separation of variables, the idea is to build the coset representatives out of the A_i. We shall only give a rough description of how this is done, together with the size of the corresponding coset spaces.

LEMMA 4.1. *The following factorizations hold: (a) $GL_n = A_2\cdots A_n\cdot P_n$. (b) $P_n = \tilde{A}_n\cdot A_{n-1}\cdots A_2\cdot A_3\cdots A_{n-1}\tilde{A}_n\cdot[GL_{n-1}\times GL_1]$ where $\tilde{A}_n = A_n\bigcap P_n$.*

These decompositions are highly redundant and do not give unique coset representatives for the coset spaces. However it is not too difficult to derive unique representatives for the cosets; see [25, 9]. Lemma 4.1 shows that we have $\gamma(GL_n > P_n, S) \le n-1$ and $\gamma(P_n > GL_n\times GL_1) \le 2n-3$. The only additional information we shall need are the sizes of the coset spaces, $|GL_n/P_n| = \frac{q^{n+1}-1}{q-1}$ and $|P_n/(GL_n\times GL_1)| = q^{n-1}$.

[2] In general, it will be useful to adopt the standard notation that if $B_1,\ldots,B_r$ are square matrices of dimensions $d_1,\ldots,d_r$, then let $\mathrm{Diag}(B_1,\ldots,B_r) = (B_1\oplus\cdots\oplus B_r)$ denote the block diagonal matrix with i^{th} block equal to B_i.

The final piece of information required by the general theory is the maximum multiplicity, $\mathcal{M}(S)$. Using a result of Thoma [35], for the restrictions of GL_n to GL_{n-1}, and some asymptotics of Stong [34] for the number of representations of GL_{n-1} we may obtain a bound on the maximum multiplicity of the restriction from GL_n to GL_{n-2} and hence that $\mathcal{M}(S) \leq 2^{2n}(q^n + K'q^{n-2})$ where K' is a constant which is independent of both n and q.

Now we have the data, a straightforward application of theorem 3.1 show us that

$$\frac{T_{GL_n}}{|GL_n|} \leq \frac{T_{P_n}}{|P_n|} + (n-1)2^{2n}(q^{2n-1} + (K'+2)q^{2n-2})$$

and

$$\frac{T_{P_n}}{|P_n|} \leq \frac{T_{GL_{n-1}}}{|GL_{n-1}|} + \frac{T_{GL_1}}{q-1} + (2n-1)2^{2n}(q^{2n-2} + K'q^{2n-4})$$

Using these relations recursively leads to a sum that can be easily evaluated to give

$$\frac{T_{GL_n}}{|GL_n|} \leq \frac{T_{GL_2}}{|GL_2|} + 3n(2q)^{2n-1} + Kn(2q)^{2n-2}.$$

There is a naive bound for T_{GL_2} of q^8 operations so by a slight alteration of the constant K, we obtain the desired result for $n \geq 3$. When $n = 2$, a generalization of these methods shows that $T_{GL_2} \leq |GL_2|\,(5q-1)$.

Remarks. 1. Variations of the algorithm. There is of course nothing canonical about either the generators chosen here for GL_n or the subgroup chain. It seems highly likely that better choices for either are possible. Always, commutativity will need to be exploited and here it may be necessary to effectively compute the centralizers of various subsets of elements. Recent advances in computational group theory for matrix groups [4] may prove useful.

In fact slightly more complicated variants of Theorem 3.2 and its Corollary can improve Theorem 4.4. For an indication of this approach see Section 6.

2. Earlier work. The problem of finding an efficient algorithm for computing a DFT for $GL_n(q)$ was first considered in [28]. There an algorithm is proposed which uses "models" (direct sums of induced one-dimensional representations which contain each irreducible of the group exactly once) to compute a DFT for GL_n. In so doing the algorithm proceeds in two parts: (1) Computing the Fourier transform at reducible representations which are given by monomial matrices and then (2) applying projection operators to these reducible matrices in order to obtain collection of unique irreducible Fourier transforms. Some simple asymptotics for the bounds they obtain yield an estimate for the complexity of their algorithm to be

$$O(|GL_n(q)|q^{\frac{n^2-2n}{4}}).$$

3. Direct approach. It is also necessary to compare our algorithm with the algorithm which uses the subgroup chain but does not factor the coset representatives and thus performs direct matrix multiplication of the twiddle factors. Straightforward analysis then shows that such an algorithm yields an upper bound which depends on the maximum degree of an irreducible representation of GL_n, which is of the order of $q^{\frac{1}{2}(n^2-n)}$. This direct algorithm gives an upper bound of

$$O(nq^{\frac{1}{2}(n^2-3n)}|GL_n(q)|).$$

4.4 Finite groups of Lie type and their generalizations. The techniques used to compute the DFT in GL_n may be extended in a relatively straightforward manner, to Chevalley groups, finite groups of Lie type, or in a more abstract setting to any finite group with a split BN pair which satisfies appropriate commutator relations. We refer the reader to the book of Carter [9] for definitions. These groups are essentially subgroups of $GL_n(q)$ - so-called finite classical groups.

Any finite group of Lie type, G, has a subgroup chain analogous to (4.15), where P_{n-1} is replaced by a maximal **parabolic subgroup**, $GL_{n-1} \times GL_1$ by its **reductive part**, and GL_{n-1} by the **semisimple part** of the parabolic subgroup. Each Lie group of finite type G has an associated **Weyl group**, which has an analogous relation to G as does the symmetric group to GL_n, where S_n is embedded in GL_n as permutation matrices. For each simple reflection in the Weyl group (in S_n, these are the pairwise adjacent transpositions) there is a corresponding subgroup of G with a "split (B, N)-pair of type A_1" which is generated by the positive and negative root subgroups of G associated to that simple reflection. These subgroups correspond to the A_i defined in the proof of Theorem 4.4 and we take their union as our generating set.

The factorization $GL_n = A_2 \cdots A_n \cdot P_n$ comes from a factorization of the minimal coset representative of S_n/S_{n-1} of maximal length, i.e. $t_2 \ldots t_n$. There is a similar relationship between any finite group of Lie type and its Weyl group; there is a factorization of cosets of G relative to a parabolic subgroup corresponding to a factorization of coset representatives in the Weyl group. Factorizing coset representatives of the parabolic subgroup relative to its reductive part simply amounts to factorizing elements of its largest normal unipotent subgroup (in GL_n this is the abelian subgroup of matrices with 1s on the diagonal, and zeroes everywhere else except for the last column), and this can be done within the Borel subgroup (in GL_n, the upper triangular matrices).

The most difficult part of the GL_n calculation to generalize to this setting is the bound on the multiplicities of the restrictions to parabolic subgroups. For GL_n, we used an explicit result of Thoma [35], but for more general finite groups of Lie type this is still a matter of current research. Recent results of Hagedorn [21], indicate that if the parabolic subgroup is well chosen then an efficient algorithm will result. For example, for the groups $B_n(q)$ which are essentially the odd special orthogonal groups of $\mathbf{F}_q$, Hagedorn has shown [21] that the maximum multiplicity of the restriction from $B_n(q)$ to $B_{n-2}(q)$ is $O(q^{3n-2})$. Using this result, the techniques of the preceding section yield a DFT of complexity $O(q^{5n-3}|B_n(q)|)$ for fixed n.

5 Proof Sketch of Main Theorem

The fundamental technique introduced in this paper is the use of commutativity properties of the group G to construct a chain of subgroups and accompanying coset representatives with advantageous computational properties. To make use of commutativity between certain elements of G, Schur's lemma is applied. There are a number of essentially equivalent formulations of Schur's Lemma. The following form is most useful for this paper.

Lemma 5.1. (Schur) *Let $K < G$, and suppose that a is in the centralizer of K. Suppose that ρ is a K-adapted representation of G so that $\rho(a) = \eta_1(a) \oplus \cdots \oplus \eta_1(a) \oplus \cdots \oplus \eta_r(a) \oplus \cdots \oplus \eta_r(a)$ where $\eta_1, \ldots, \eta_r$ inequivalent irreducible matrix representations of K, and η_i occurs with multiplicity m_i. Then up to a permutation of rows and columns, $\rho(a)$ is of the form*

$$(GL_{m_1}(\mathbf{C}) \otimes I_{d_1}) \oplus \cdots (GL_{m_r}(\mathbf{C}) \otimes I_{d_r}) \tag{5.17}$$

where I_k denotes the $k \times k$ identity matrix, $\otimes$ the usual tensor product of matrices, and $d_i = d_{\eta_i}$ is the dimension of η_i

From a computational point of view (5.17) expresses the fact that commutativity combined with adaptability imply that the commuting matrix will be sparse.

Now suppose that $G \geq H \geq K$ and a commutes with K and is also contained in H. If ρ is adapted to both H and K, then $\rho(a)$ will not only have the form (5.17), but also have block diagonal form according to the decomposition of the restriction of ρ to H. A simple count of the number and position of the nonzero entries of $\rho(a)$ then gives us the following theorem which immediately implies Lemma 3.1.

Theorem 5.1. *Assume $G > H > K$ is a chain of subgroups of G, and ρ is a chain-adapted representation of G. Let $\{\mu_i\}$, $\{\eta_j\}$ be complete sets of representations of H and K respectively. Let m_i denote the multiplicity of μ_i in the restriction ρ, and let m_{ij} denote the multiplicity η_j in the restriction of μ_j. Then the matrix multiplication $\rho(a) \cdot F$ can be performed for any $d_\rho \times d_\rho$ matrix F performed in*

$$\sum_{i,j} m_i m_{ij}^2 d_j d_\rho \leq \mathcal{M}(H,K) d_\rho^2 \tag{5.18}$$

operations.

6 Further improvements and directions

1. Variations on the main results. Theorem 3.1 and its corollary are particularly easy to use but are by no means the best results possible. A more careful count of the number of operations for the algorithms corresponding to these theorems gives an immediate improvement; complicated expressions but exact expressions for the number of operations can be obtained, but they all essentially derive from the ideas of Sections 3 and 5.

Even further improvements are possible. By working on the level of matrix entries rather than the matrices themselves we may perform the matrix products occuring in the above results in any order; this requires a much more complicated indexing scheme and the partial results are no longer matrices but more general indexed quantities. It is possible to give an explicit expression for the complexity of the resulting algorithms, in a form similar to but generalizing (3.1). For certain groups these algorithms are more efficient than the ones above. As an example, we get a better bound on the complexity of the DFT on $GL_n(q)$ using the same bases as in section 4.

Theorem 6.1. *For any n, there is a positive contant, K_n, such that*

$$T_{GL_n(q)} \leq K_n q^n |GL_n(q)|$$

for any $q \geq 2$.

2. Homogeneous spaces. For many statistical applications data on homogeneous spaces is of interest, rather than data on the full group. In brief, a homogeneous space for a finite group is simply a set on which the group acts transitively as permutations. A common example is the action of the finite affine group on point-line pairs and more generally, the action of an automorphism group of a design on its block-point pairs. In this case generalizations of the "usual" analysis of variance for data on such sets requires the computation of projections of the data vector onto group-invariant subspaces. The ideas of Section 3 may be applied and again speed-ups of the currently known most efficient algorithms (cf. [17] and references therein) can be obtained.

References

[1] L. Babai, E Luks, and A. Seress. Fast management of permutation groups, *Proc.* 28^{th} *IEEE FOCS* (1988), pp. 272-282.

[2] L. Babai and L. Rónyai. Computing irreducible representations of finite groups. *Math. Comp.* **55** (1990), 705-722.

[3] L. Babai, K. Freidl and M. Stricker. Decomposition of *-closed algebras in polynomial time. *Proc. of* 18^{th} *ACM ISSAC* (1993), pp. 86-94.

[4] R. Beals. and L. Babai. Las Vegas algorithms for matrix groups. *Proc.* 34^{th} *IEEE FOCS* (1993), pp. 427-436.

[5] U. Baum. Existence and efficient construction of fast Fourier transforms for supersolvable groups. *Computational Complexity*, **1** (1991), 235-256.

[6] U. Baum and M. Clausen. Some lower and upper complexity bounds for generalized Fourier transforms and their inverses. *SIAM J. Comput.* **20**(3) (1991), 451-459.

[7] T. Beth. On the computational complexity of the general discrete Fourier transform, *Theoretical Computer Science* **51** (1987), 331-339.

[8] N. Bshouty, M. Kaminski, and D. Kirkpatrick. Addition requirements for matrix and transposed matrix products. *J. of Algorithms* **9** (1988), 354-364.

[9] R. Carter. *Simple Groups of Lie Type.* Wiley-Interscience, NY (1989).

[10] M. Clausen and U. Baum. *Fast Fourier Transforms*, Wissenschaftsverlag, Manheim (1993).

[11] M. Clausen and U. Baum, Fast Fourier transforms for symmetric groups, theory and implementation. *Math. Comp.* **61**(204) (1993), 833-847.

[12] M. Clausen. Fast Fourier transforms metabelian groups. *SIAM J. Comput.* **18** (1989), 584-593.

[13] M. Clausen. Fast generalized Fourier transforms. *Theor. Comp. Sci.* **67** (1989), 55-63.

[14] J. W. Cooley and J. W. Tukey. An algorithm for machine calculation of complex Fourier series. *Math. Comput.* **19** (1965), 297-301.

[15] P. Diaconis. *Group Representations in Probability and Statistics*, IMS, Hayward, CA (1988).

[16] P. Diaconis and D. Rockmore. Efficient computation of the Fourier transform on finite groups. *J. of the A.M.S.* **3**(2) (1990), 297-332.

[17] P. Diaconis and D. Rockmore. Efficient computation of isotypic projections for the symmetric group. *DIMACS Series in Disc. Math., Vol. 11* (eds. L. Finkelstein and W. Kantor), 87-104 (1993).

[18] J. R. Driscoll and D. Healy. Computing Fourier transforms and convolutions on the 2-sphere. (Extended abstract) *Proc.* 34^{th} *IEEE FOCS*, (1989) pp. 344-349; (*Adv. in Appl. Math.*, To appear.

[19] D. F. Elliott and K. R. Rao. *Fast Transforms: Algorithms, Analyses, and Applications.* Academic, New York (1982).

[20] L. Finkelstein and W. Kantor. *Groups and Computation*, DIMACS Series in Disc. Math., Vol. 11, AMS (1993).

[21] T. Hagedorn. *Multiplicities in Restricted Representations of* $GL_n(\mathbf{F}_q)$, $U_n(\mathbf{F}_q)$ *and* $SO_n(\mathbf{F}_q)$. Ph.D. Thesis, Department of Mathematics, Harvard University (1994).

[22] F. Harary. *Graph Theory.* Addison-Wesley, Reading, MA (1972).

[23] D. Healy, D. Maslen, S. Moore, and D. Rockmore. Applications of a fast convolution algorithm on the 2-sphere. *Technical Report, Department of Mathematics and Computer Science, Dartmouth College* (1994).

[24] G. D. James. *The Representation Theory of the Symmetric Groups.* Lecture Notes in Mathematics., Vol. 682, Springer-Verlag, Berlin (1978).

[25] G. D. James. *Representations of General Linear Groups.* LMS Vol. 94, Cambridge Univ. Press, Cambridge (1984).

[26] M. Karpovsky and E. Trachtenberg, Filtering in a communication channel by Fourier transforms over finite groups, in *Spectral Techniques and Fault Detection*, M. Karpovsky (ed.) Acaddemic Press, NY (1985), pp. 179-212.

[27] A. Kerber. *Representations of Permutations Groups I, II* Lecture Notes in Mathematics, Vols. 240 and 495, Springer-Verlag, Berlin (1971) and (1975).

[28] S. Linton, G. Michler, and J. Olsson. Fourier transforms with respect to monomial representations. *Math. Ann.* **297** (1993), 253-268.

[29] D. Rockmore. Efficient computation of Fourier inversion for finite groups. *J. of the A.C.M.* **41**(1) (1994), 31-66.

[30] D. Rockmore. Fast Fourier analysis for abelian group extensions. *Adv. in Appl. Math.* **11** (1990), 164-204.

[31] D. Rockmore. Fast Fourier transforms for wreath products. *Technical Report, Dept. of Mathematics, Dartmouth College* (1994).

[32] J. P. Serre. *Linear Representations of Finite Groups.* Springer-Verlag, New York (1977).

[33] C. C. Sims, Computational methods in the study of permutation groups, *in: Computational Problems in Abstract Algebra*, J. Leech, ed., Pergamon Press 1970, pp. 169-183.

[34] R. Stong. Some asymptotic results on finite vector spaces. *Adv. in Appl. Math.* **9** (1988), 167-199.

[35] E. Thoma. Die Einschrankung der Charaktere von $GL(n,q)$ auf $GL(n-1,q)$. *Math. Zeit.*, **119** (1971) 321-338.

[36] R. Tolimieri, M. An, and C. Lu. *Algorithms for Discrete Fourier Transform and Convolution.* Springer-Verlag, New York (1989).

[37] C. Van Loan. *Computational Framework for the Fast Fourier Transform*, SIAM, Philadelphia (1992).