SIAM J. COMPUT.
Vol. 37, No. 6, pp. 1842–1864

THE SYMMETRIC GROUP DEFIES STRONG FOURIER SAMPLING*

CRISTOPHER MOORE†, ALEXANDER RUSSELL‡, AND LEONARD J. SCHULMAN§

Abstract. The dramatic exponential speedups of quantum algorithms over their best existing classical counterparts were ushered in by the technique of *Fourier sampling*, introduced by Bernstein and Vazirani and developed by Simon and Shor into an approach to the hidden subgroup problem. This approach has proved successful for abelian groups, leading to efficient algorithms for factoring, extracting discrete logarithms, and other number-theoretic problems. We show, however, that this method cannot resolve the hidden subgroup problem in the symmetric groups, even in the weakest, information-theoretic sense. In particular, we show that the GRAPH ISOMORPHISM problem cannot be solved by this approach. Our work implies that any quantum approach based upon the measurement of coset states must depart from the original framework by using entangled measurements on multiple coset states.

Key words. quantum computing, hidden subgroup problem, graph isomorphism, Fourier sampling

AMS subject classifications. 68Q17, 81R05, 43A65

DOI. 10.1137/050644896

1. Introduction: The hidden subgroup problem.

Many problems of interest in quantum computing can be reduced to an instance of the *hidden subgroup problem* (HSP). We are given a group G and a function f with the promise that, for some subgroup $H \subseteq G$, f is invariant precisely under translation by H; that is, f is constant on the left cosets of H and takes distinct values on distinct cosets. We then wish to determine the subgroup H by querying f. Most algorithms for the HSP use the following approach, referred to as the *standard method* or *Fourier sampling* [5].

Step 1. Prepare two registers, the first in a uniform superposition over the elements of G and the second with the value zero, yielding the state

$$|\psi_1\rangle = \frac{1}{\sqrt{|G|}} \sum_{g \in G} |g\rangle \otimes |0\rangle \quad .$$

Step 2. Query (or calculate) the function f defined on G and XOR it with the second register. This entangles the two registers and results in the state

$$|\psi_2\rangle = \frac{1}{\sqrt{|G|}} \sum_{g \in G} |g\rangle \otimes |f(g)\rangle \quad .$$

Step 3. Measure the second register. This puts the first register in a uniform superposition over one of f's level sets, i.e., one of the left cosets of H, and

*Received by the editors November 11, 2005; accepted for publication (in revised form) November 12, 2007; published electronically March 26, 2008. This work was supported by NSF grants CCR-0093065, PHY-0200909, PHY-0456720, EIA-0218443, EIA-0218563, CCR-0220070, CCR-0220264, CCF-0524828, and CCF-0524613, and ARO grants W911NF-04-R-0009 and W911NF-05-1-0294.
http://www.siam.org/journals/sicomp/37-6/64489.html

†Department of Computer Science, University of New Mexico, Albuquerque, NM 87131, and the Santa Fe Institute, Santa Fe, NM 87501 (moore@cs.unm.edu).

‡Department of Computer Science and Engineering, University of Connecticut, Storrs, CT 06269 (acr@cse.uconn.edu).

§Computer Science Department, California Institute of Technology, Pasadena, CA 91125 (schulman@cs.caltech.edu).

disentangles it from the second register. If we observe the value $f(c)$, we have the state $\psi_3 \otimes |f(c)\rangle$, where

$$|\psi_3\rangle = |cH\rangle = \frac{1}{\sqrt{|H|}} \sum_{h \in H} |ch\rangle \quad .$$

Alternately, we can view the first register as being in a mixed state with density matrix

$$\rho = \frac{1}{|G|} \sum_{g \in G} |cH\rangle \langle cH| \quad .$$

Step 4. Carry out the quantum Fourier transform on $|\psi_3\rangle$ and measure the result; that is, observe the "frequency" corresponding to one of the Fourier basis functions.

For example, in Simon's problem [35], $G = \mathbb{Z}_2^n$ and f is an oracle such that, for some y, $f(x) = f(x+y)$ for all x; in this case $H = \{0, y\}$ and we wish to identify y. In Shor's factoring algorithm [34], G is essentially the group $\mathbb{Z}_n^*$, where n is the number we wish to factor, $f(x) = c^x \bmod n$ for a random $c < n$, and H is the subgroup of $\mathbb{Z}_n^*$ whose index is the multiplicative order of c. (However, Shor's algorithm does not operate on $\mathbb{Z}_n^*$ directly—indeed, knowing $|\mathbb{Z}_n^*|$ would provide an efficient classical algorithm. Instead, it performs the quantum Fourier transform over $\mathbb{Z}_q$ for some $q = \text{poly}(n)$; see [34] or [13, 14].)

In both Simon's and Shor's algorithms, the group G is abelian and finite. It is not hard to see that, in this case, a polynomial number (i.e., polynomial in $\log |G|$) of experiments of this type determine H. In a cyclic group, for instance, the observed frequency is a random multiple of the index of H, so we can determine this index with high probability by taking the greatest common divisor of these frequencies. More generally, each experiment yields a random element of the dual space $H^\perp$ perpendicular to H's characteristic function. After $O(\log |G|)$ such experiments, with high probability these elements span $H^\perp$, and we can determine H via linear algebra.

While the *nonabelian* HSP appears to be much more difficult, it has very attractive applications. In particular, solving the HSP for the symmetric group S_n would provide an efficient quantum algorithm for the Graph Automorphism and Graph Isomorphism problems (see, e.g., Jozsa [21] for a review). Another important motivation is the relationship between the HSP over the dihedral group with hidden shift problems [7] and cryptographically important cases of the shortest lattice vector problem [29].

So far, algorithms for the HSP are known for only a few families of nonabelian groups, including groups whose commutator subgroup is of polynomial size [30, 20]; "smoothly solvable" groups [10]; and some semidirect products of abelian groups [28, 18, 3]. Ettinger and Høyer [8] provided another type of result by showing that Fourier sampling can solve the HSP for the dihedral groups D_n in an *information-theoretic* sense. That is, a polynomial number of experiments gives enough information to reconstruct the subgroup, though it is unfortunately not known how to determine H from this information in polynomial time.

Extending the notion of Fourier sampling to nonabelian groups requires that we define a nonabelian version of the Fourier transform. For abelian groups, the Fourier basis functions are simply the homomorphisms $\phi : G \to \mathbb{C}$ such as the familiar exponential function $\phi_k(x) = e^{2\pi i k x/n}$ for the cyclic group $\mathbb{Z}_n$. In the nonabelian case,

there are not enough such homomorphisms to provide a basis for all $\mathbb{C}$-valued functions on G. To create such a basis, we generalize to the *representations* of the group, namely, homomorphisms $\rho : G \to \mathsf{U}(V)$, where $\mathsf{U}(V)$ is the group of unitary matrices acting on some $\mathbb{C}$-vector space V of dimension d_ρ. It suffices to consider *irreducible* representations, namely, those for which no nontrivial subspace of V is fixed by the various operators $\rho(g)$; Fourier analysis over abelian groups then corresponds to the special case where all irreducible representations have dimension one, the single entry in these 1×1 matrices being the values of the Fourier basis functions. In general, once a basis for each irreducible ρ is chosen, the matrix elements ρ_{ij} provide an orthogonal basis for the vector space of all $\mathbb{C}$-valued functions on G.

The quantum Fourier transform then consists of transforming (unit-length) vectors in $\mathbb{C}[G] = \{\sum_{g\in G} \alpha_g |g\rangle \mid \alpha_g \in \mathbb{C}\}$ from the basis $\{|g\rangle \mid g \in G\}$ to the basis $\{|\rho, i, j\rangle\}$, where ρ is the name of an irreducible representation and $1 \le i, j \le d_\rho$ index a row and a column (in a chosen basis for V). Note, however, that a representation $\rho : G \to \mathsf{U}(V)$ does not intrinsically distinguish any specific basis for the underlying space V and, for high-dimensional representations, this appears to require a rather dramatic choice on the part of the transform designer. For instance, in a group such as S_n, in most bases a typical representation $\rho(g)$ is a dense matrix of exponential size, but for a carefully chosen basis it is sparse and highly structured. Making such choices of bases allows us to efficiently carry out the quantum Fourier transform for a wide variety of groups [4, 17, 27].

Since the work of [15, 12], the most fundamental question concerning the HSP has been whether there is a basis for the irreducible representations of a given group such that measuring coset states in this basis provides enough information to determine H and, if so, whether this information can be extracted by an efficient algorithm. This framework is known as *strong Fourier sampling*. In this article, we answer this question in the negative for the symmetric group S_n, showing that this process cannot distinguish relevant subgroups from each other, or from the trivial subgroup, even information-theoretically. Indeed, we show that no measurement whatsoever, including arbitrary positive operator-valued measurements (POVMs), on single coset states can succeed. We remark that the subgroups on which we focus are among the most important special cases of the HSP, as they are those to which Graph Isomorphism naturally reduces.

Related work. The terminology "strong Fourier sampling" [12] was invented to distinguish this approach from the natural variant, called *weak Fourier sampling*, where one only measures the name of the representation ρ and ignores the row and column information. Weak Fourier sampling is basis-independent, making it attractive from the standpoint of analysis; however, it cannot distinguish conjugate subgroups from each other, and Hallgren, Russell, and Ta-Shma [15] showed that it cannot distinguish the trivial subgroup from an order-2 subgroup of S_n consisting of $n/2$ disjoint transpositions. Specifically, they used character bounds to show that the probability distributions obtained on representation names for the trivial and order-2 subgroups are exponentially close in total variation distance: thus one needs an exponential number of such experiments to distinguish them. Kempe and Shalev [22] have generalized this result to other conjugacy classes and conjectured that one can do no better than classical computation with this approach.

In an effort to shed light on the power of strong Fourier sampling, Grigni et al. [12] showed that, for groups such as S_n, measuring in a *random* basis yields an exponentially small amount of information. This can be explained, roughly, by the fact that projecting a vector into a sufficiently high-dimensional random subspace results in

tightly concentrated length. On the other hand, Moore et al. [28] showed that for the affine groups and some q-hedral groups, measuring in a well-chosen basis can solve the HSP in cases where random bases cannot.

Our contribution. In this paper we show that strong Fourier sampling, in an arbitrary basis of the algorithm designer's choice, cannot solve the HSP for S_n. As in [15] we focus on order-2 subgroups of the form $\{1, m\}$, where m is an involution consisting of $n/2$ disjoint transpositions. We show that strong Fourier sampling—and more generally, arbitrary measurements of single coset states—cannot distinguish most subgroups of this form from each other, or from the trivial subgroup, without an exponential number of experiments.

The motivation for looking at this case of the HSP is as follows. If we fix two rigid connected graphs of size n, then the automorphism group H of their disjoint union is a subgroup of S_{2n}. If they are isomorphic, then H is of the form $\{1, m\}$, where m is the involution that swaps the two graphs, while if they are nonisomorphic, then H is trivial. This yields a classical reduction from Graph Isomorphism to Graph Automorphism, and our results preclude a quantum algorithm for the latter problem that works by reducing to the HSP on the symmetric group.

However, the involutions m which switch the two graphs are not generic elements of the conjugacy class in S_{2n} consisting of n disjoint transpositions, since they switch the first n vertices with the last n vertices. The set of such elements forms a conjugacy class in the *wreath product* $S_n \wr \mathbb{Z}_2 \subset S_{2n}$, and it is the HSP on this group, rather than all of S_{2n}, to which Graph Isomorphism naturally reduces. To address the possibility of a quantum algorithm that uses this reduction, we present an additional result showing that this case of the HSP also requires an exponential number of experiments.

We remark that our results do not preclude the existence of an efficient quantum algorithm for the HSP on S_n or $S_n \wr \mathbb{Z}_2$. Rather, they force us to either abandon coset states or consider *multiregister* algorithms, in which we prepare multiple coset states and subject them to entangled measurements, rather than performing a product measurement where each coset state is measured independently. Some progress in this direction has been made: Ettinger, Høyer, and Knill [9] showed that the HSP on arbitrary groups can be solved information-theoretically with a polynomial number of coset states, and two of the present authors have shown how to carry out such a measurement in the Fourier basis [25]. Kuperberg [24] devised a subexponential $(2^{O(\sqrt{\log n})})$ algorithm for the HSP on the dihedral group D_n that uses entangled measurements, and Alagic, Moore, and Russell [1] obtained a similar algorithm for the HSP on groups of the form G^n. Bacon, Childs, and van Dam determined the *optimal* multiregister measurement for the dihedral group [2] (see also [26]) and used this approach to construct an algorithm for a class of semidirect product groups [3].

Whether a similar approach can be applied to the symmetric group is a major open question. Hallgren et al. [16] have shown, however, that no family of measurements across $o(n \log n)$ coset states can distinguish $H = \{1, m\}$ from the trivial group in S_n with a polynomial number of repetitions. We remark that in light of the upper bounds of [9, 25], $O(n \log n)$ coset states do, at least information-theoretically, determine the answer. Constructing such highly entangled measurements poses a major conceptual challenge, and it is far from clear in what cases they can be carried out efficiently. Of course, it is also possible that a completely different approach—one which does not use coset states, or which does not start by reducing to the HSP—will provide an efficient quantum algorithm for Graph Isomorphism.

The paper is organized as follows. In section 2 we give a brief introduction to

representation theory and nonabelian Fourier analysis. In section 3 we discuss the general structure of quantum measurements on coset states and show that the optimal measurement takes the form of strong Fourier sampling. In section 4 we show how to bound the variance of the resulting probability distributions with respect to the choice of hidden subgroup. In section 5 we record some specific facts about the representations of the symmetric group, and in section 6 we use these facts to show that an exponential number of measurements are necessary. Finally, in section 7 we adapt the argument for the specific family of involutions relevant to Graph Isomorphism.

2. Fourier analysis over finite groups. We briefly discuss the elements of the representation theory of finite groups. Our treatment is primarily for the purposes of setting down notation; we refer the reader to [11, 33] for complete accounts.

Let G be a finite group. A *representation* ρ of G is a homomorphism $\rho : G \to \mathsf{U}(V)$, where V is a finite-dimensional Hilbert space and $\mathsf{U}(V)$ is the group of unitary operators on V. The *dimension* of ρ, denoted d_ρ, is the dimension of the vector space V. By choosing a basis for V, we can then identify $\rho(g)$ with a unitary $d_\rho \times d_\rho$ matrix so that for every $g, h \in G$, $\rho(gh) = \rho(g) \cdot \rho(h)$.

Fixing a representation $\rho : G \to \mathsf{U}(V)$, we say that a subspace $W \subset V$ is *invariant* if $\rho(g)W \subset W$ for all $g \in G$. We say ρ is *irreducible* if it has no invariant subspaces other than the trivial space $\{\mathbf{0}\}$ and V. If two representations ρ and σ are the same up to a unitary change of basis, we say that they are *equivalent*. It is a fact that any finite group G has a finite number of distinct irreducible representations up to equivalence, and, for a group G, we let $\widehat{G}$ denote a set of representations containing exactly one from each equivalence class. The irreducible representations of G give rise to the Fourier transform. Specifically, for a function $f : G \to \mathbb{C}$ and an element $\rho \in \widehat{G}$, define the *Fourier transform of f at ρ* to be

$$\hat{f}(\rho) = \sqrt{\frac{d_\rho}{|G|}} \sum_{g \in G} f(g)\rho(g) \ .$$

The leading coefficients are chosen to make the transform unitary, so that it preserves inner products:

$$\langle f_1, f_2 \rangle = \sum_g f_1^*(g) f_2(g) = \sum_{\rho \in \widehat{G}} \mathbf{tr}\left(\hat{f}_1(\rho)^\dagger \cdot \hat{f}_2(\rho)\right) \ .$$

Given a representation ρ and pair of integers $1 \le i, j \le d_\rho$, we can associate a basis vector $|\rho, i, j\rangle$, which assigns the matrix entry $\rho(g)_{i,j}$ to each element g. As described above, these form an orthonormal basis for $\mathbb{C}[G]$, which implies

$$\sum_{\rho \in \widehat{G}} d_\rho^2 = |G| \ .$$

In the case when ρ is *not* irreducible, it can be decomposed into a direct sum of irreducible representations, each of which operates on an invariant subspace. We write $\rho = \sigma_1 \oplus \cdots \oplus \sigma_k$ and, for the σ_i appearing at least once in this decomposition, $\sigma_i \prec \rho$. In general, a given σ can appear multiple times, in the sense that ρ can have an invariant subspace isomorphic to the direct sum of a_σ^ρ copies of σ. In this case a_σ^ρ is called the *multiplicity* of σ in ρ, and we write $\rho = \bigoplus_{\sigma \prec \rho} a_\sigma^\rho \sigma$.

For a representation ρ we define its *character* as the trace $\chi_\rho(g) = \mathbf{tr}\,\rho(g)$. Since the trace is invariant under conjugation, characters are constant on the conjugacy

classes, and if $\mathfrak{m}$ is a conjugacy class, we write $\chi_\rho(\mathfrak{m}) = \chi_\rho(m)$, where m is any element of $\mathfrak{m}$. Characters are a powerful tool for reasoning about the decomposition of reducible representations. In particular, for $\rho, \sigma \in \widehat{G}$, we have the orthogonality conditions

$$\langle \chi_\rho, \chi_\sigma \rangle_G = \frac{1}{|G|} \sum_{g \in G} \chi_\rho(g) \chi_\sigma(g)^* = \begin{cases} 1, & \rho = \sigma \ , \\ 0, & \rho \neq \sigma \ . \end{cases}$$

If ρ is reducible, we have $\chi_\rho = \sum_{\sigma \prec \rho} a_\sigma^\rho \chi_{\sigma_i}$, and so the multiplicity a_σ^ρ is given by

$$a_\sigma^\rho = \langle \chi_\rho, \chi_\sigma \rangle_G \ .$$

If ρ is irreducible, *Schur's lemma* asserts that the only matrices which commute with $\rho(g)$ for all g are the scalars, $\{c\mathbb{1} \mid c \in \mathbb{C}\}$. Therefore, for any A we have

$$\frac{1}{|G|} \sum_{g \in G} \rho(g)^\dagger A \rho(g) = \frac{\mathbf{tr}\, A}{d_\rho} \mathbb{1}_{d_\rho} \tag{2.1}$$

since conjugating this sum by $\rho(g)$ simply permutes its terms. In particular, consider the average of ρ over a conjugacy class $\mathfrak{m}$, which we denote $\rho(\mathfrak{m})$:

$$\rho(\mathfrak{m}) = \mathrm{Exp}_{m \in \mathfrak{m}}\, \rho(m) = \mathrm{Exp}_g\, \rho(g^{-1} m g) = \mathrm{Exp}_g\, \rho(g)^\dagger \rho(m) \rho(g) \ .$$

Then since $\mathbf{tr}\, \rho(m) = \chi_\rho(\mathfrak{m})$, we have

$$\rho(\mathfrak{m}) = \frac{\chi(\mathfrak{m})}{d_\rho} \mathbb{1}_{d_\rho} \ . \tag{2.2}$$

Similarly, if ρ is reducible, $\rho(\mathfrak{m})$ is scalar in each irreducible subspace, giving

$$\rho(\mathfrak{m}) = \sum_{\sigma \prec \rho} \frac{\chi_\sigma(\mathfrak{m})}{d_\sigma} \Pi_\sigma^\rho \ , \tag{2.3}$$

where Π_σ^ρ projects onto the subspace $a_\sigma^\rho \sigma$ spanned by copies of σ. We use these facts below.

There is a natural product operation on representations: if $\rho : G \to \mathsf{U}(V)$ and $\sigma : G \to \mathsf{U}(W)$ are representations of G, we may define a new representation $\rho \otimes \sigma : G \to \mathsf{U}(V \otimes W)$ by extending the rule $(\rho \otimes \sigma)(g) : \mathbf{u} \otimes \mathbf{v} \mapsto \rho(g)\mathbf{u} \otimes \sigma(g)\mathbf{v}$. In general, the representation $\rho \otimes \sigma$ is not irreducible, even when both ρ and σ are. This leads to the *Clebsch–Gordan problem*, that of decomposing $\rho \otimes \sigma$ into irreducibles. For example, since $\chi_{\rho \otimes \sigma}(g) = \chi_\rho(g) \cdot \chi_\sigma(g)$, the multiplicity of τ in $\rho \otimes \sigma$ is $\langle \chi_\tau, \chi_\rho \chi_\sigma \rangle_G$.

Group elements can act on each other on the left or right. Thus we can consider subspaces of $\mathbb{C}[G]$ that are invariant under left multiplication, right multiplication, or both; these subspaces are called *left-*, *right-*, or *bi-invariant*, respectively. Each $\rho \in \widehat{G}$ corresponds to a d_ρ^2-dimensional bi-invariant subspace of $\mathbb{C}[G]$. We can think of the bi-invariant subspace as a single d_ρ^2-dimensional representation, consisting of the space of $d_\rho \times d_\rho$ matrices A. If $\rho(g)$ acts on A by left or right multiplication, the left- and right-invariant subspaces correspond to A's columns and rows, respectively; for instance, each column of A is acted on independently by left multiplication by $\rho(g)$, and the space of matrices A which are nonzero only in this column form a d_ρ-dimensional left-invariant subspace. Thus, each bi-invariant subspace can be decomposed into

d_ρ d_ρ-dimensional left-invariant subspaces, or (transversely) d_ρ d_ρ-dimensional right-invariant subspaces.

However, this decomposition is not unique. If we think of A as the space of linear operators on the same d_ρ-dimensional vector space V on which ρ acts, changing the orthonormal basis for V transforms the matrices A. Thus, each orthonormal basis B of V gives a way to divide the bi-invariant subspace into left-invariant columns and right-invariant rows, and each such subspace is associated with some basis vector $\mathbf{b} \in B$.

3. The structure of the optimal measurement. In this section we show that starting with a single coset state, the optimal measurement for the HSP is precisely an instance of strong Fourier sampling (possibly in an overcomplete basis). This has been pointed out several times in the past, at varying levels of explicitness [19, 24]; we state it here for completeness. Everything we say in this section is true for the HSP in general. However, for simplicity we focus on the special case of the HSP called the *hidden conjugate problem* in [28]: there is a (nonnormal) subgroup H, and we are promised that the hidden subgroup is one of its conjugates, $H^g = g^{-1}Hg$ for some $g \in G$.

We may treat the states arising after Step 3 of the procedure above as elements of the group algebra $\mathbb{C}[G]$. We use the notation $|g\rangle = 1 \cdot g \in \mathbb{C}[G]$ so that the vectors $|g\rangle$ form an orthonormal basis for $\mathbb{C}[G]$. Given a set $S \subset G$, $|S\rangle$ denotes a uniform superposition over the elements of S, $|S\rangle = (1/\sqrt{|S|}) \sum_{s \in S} |s\rangle$.

3.1. The optimal POVM consists of strong Fourier sampling. The most general type of measurement allowed in quantum mechanics is a POVM. A POVM with a set of possible outcomes J consists of a set of positive operators $\{M_j \mid j \in J\}$ subject to the completeness condition,

$$\sum_j M_j = \mathbb{1} \ . \tag{3.1}$$

Since positive operators are self-adjoint, they can be orthogonally diagonalized, and since their eigenvalues are positive, they can be written as a positive linear combination of projection operators (see, e.g., [32, sect. 10]). Any POVM may thus be refined so that each $M_j = a_j \mu_j$, where μ_j is a projection operator and a_j is positive and real.

The result of this refined measurement on the state $|\psi\rangle$ is a random variable, taking values in J, that is equal to $j \in J$ with probability $P_j = a_j \langle\psi| \mu_j |\psi\rangle$. Note that the outcomes j need not correspond to subgroups directly; the algorithm designer is free to carry out t of these experiments (where t is, ideally, polynomial), observing outcomes $j_1, \ldots, j_t$, and then apply some additional computation to find the most likely subgroup given these observations.

If g is chosen from G uniformly so that the hidden subgroup is a uniformly random conjugate of H, we wish to find a POVM that maximizes the probability of correctly identifying g from the coset state $|H^g\rangle$. (Of course, to identify a conjugate H^g, we need only specify g up to an element of the normalizer of H.) Since a random left coset of H^g can be written $cgH^g = cHg$ for a random $c \in G$, the probability we observe outcome j is

$$P_j = a_j \frac{1}{|G|} \sum_{c \in G} \langle cHg| \mu_j |cHg\rangle \ . \tag{3.2}$$

Ip [19] observed that in the special case that each outcome j corresponds to a subgroup, maximizing the probability that j is correct subject to the constraint (3.1)

gives a semidefinite program. Since such programs are convex, the optimum is unique and is a fixed point of any symmetries possessed by the problem.

However, our proof relies on an elementary "symmetrization" argument. Given a group element $x \in G$, let $L_x |g\rangle = |xg\rangle$ denote the unitary matrix corresponding to left group multiplication by x. In particular, applying L_x maps one left coset onto another: $|cHg\rangle = L_c |Hg\rangle$. Writing

$$P_j = a_j \frac{1}{|G|} \sum_{c \in G} \langle cHg | \mu_j | cHg \rangle = a_j \left\langle Hg \left| \frac{1}{|G|} \sum_{c \in G} L_c^\dagger \mu_j L_c \right| Hg \right\rangle \ ,$$

we conclude that replacing μ_j for each j with the symmetrization

$$\mu_j' = \frac{1}{|G|} \sum_{g \in G} L_g^\dagger \mu_j L_g$$

does not change the resulting probability distribution P_j. Since μ_j' commutes with L_x for every $x \in G$ and provides exactly the same information as the original μ_j, we may assume without loss of generality that the optimal POVM commutes with L_x for every $x \in G$.

It is easy to see that any projection operator that commutes with left multiplication projects onto a left-invariant subspace of $\mathbb{C}[G]$, and we can further refine the POVM so that each μ_j projects onto an *irreducible* left-invariant subspace. Each such space is contained in the bi-invariant subspace corresponding to some irreducible representation ρ, in which case we write $\mathbf{im}\, \mu_j \subseteq \rho$. As discussed in section 2, a given irreducible left-invariant subspace corresponds to some unit vector $\mathbf{b}$ in the vector space V on which ρ acts. Thus we can write

$$\mu_j = |\mathbf{b}_j\rangle \langle \mathbf{b}_j| \otimes \mathbb{1}_{d_\rho} \ ,$$

where $\mathbb{1}_{d_\rho}$ is the identity operator on that left-invariant subspace. Let $B = \{\mathbf{b}_j \mid \mathbf{im}\, \mu_j \in \rho\}$; then (3.1) implies a completeness condition for each $\rho \in \widehat{G}$,

$$\sum_{\mathbf{b}_j \in B} a_j |\mathbf{b}_j\rangle \langle \mathbf{b}_j| = \mathbb{1}_{d_\rho} \ , \tag{3.3}$$

and so B is a (possibly overcomplete) basis for V. In other words, the optimal POVM consists of first measuring the representation name ρ and then performing a POVM on the vector space V with the set of possible outcomes B. Another way to see this is to regard the choice of coset as a mixed state; then its density matrix is block-diagonal in the Fourier basis, and so as Kuperberg puts it [24], measuring the representation name "sacrifices no entropy."

We note that in the special case that this POVM is a von Neumann measurement—that is, when B is an orthonormal basis for V—it corresponds to measuring the column of ρ in that basis, which is how strong Fourier sampling is usually defined. (As pointed out in [12], nothing is gained by measuring the row, since we have a random left coset cHg and left-multiplying by a random element c in an irreducible representation completely mixes the probability across the rows in each column. Here this is reflected by the fact that each μ_j is a scalar in its left-invariant subspace.) However, in general the optimal measurement might consist of an overcomplete basis, or *frame*, in each ρ, consisting of vectors $\mathbf{b}_j$ with weights a_j.

Now that we know μ_j takes this form, let us change notation. Given $\rho \in \widehat{G}$ acting on a vector space V and a unit vector $\mathbf{b} \in V$, let $\Pi^\rho_{\mathbf{b}} = |\mathbf{b}\rangle\langle\mathbf{b}| \otimes \mathbb{1}_{d_\rho}$ denote the projection operator onto the left-invariant subspace corresponding to $\mathbf{b}$. Then $\mu_j = \Pi^\rho_{\mathbf{b}_j}$, and (3.2) becomes

$$P_j = a_j \frac{1}{|G|} \sum_{c \in G} \left\| \Pi^\rho_{\mathbf{b}_j} |cHg\rangle \right\|^2 = a_j \left\| \Pi^\rho_{\mathbf{b}_j} |Hg\rangle \right\|^2 \quad . \tag{3.4}$$

We can write this as the product of the probability $P(\rho)$ that we observe ρ, times the conditional probability $P(\rho, \mathbf{b}_j)$ that we observe $\mathbf{b}_j$. Note that by (3.3),

$$\Pi^\rho = \sum_{\mathbf{b}_j \in B} a_j \Pi^\rho_{\mathbf{b}_j}$$

is the projection operator onto the bi-invariant subspace corresponding to ρ. Then

$$P_j = P(\rho) P(\rho, \mathbf{b}_j) \ ,$$

where

$$P(\rho) = \|\Pi^\rho |H\rangle\|^2 \ , \tag{3.5}$$

$$P(\rho, \mathbf{b}_j) = a_j \left\| \Pi^\rho_{\mathbf{b}_j} |Hg\rangle \right\|^2 / P(\rho) \ . \tag{3.6}$$

Note that $P(\rho, \mathbf{b}_j)$ depends on g, but $P(\rho)$ does not, which is why weak sampling is incapable of distinguishing conjugate subgroups.

3.2. The probability distribution for a conjugate subgroup. Now let us use the fact that $|H\rangle$ is a superposition over a subgroup and calculate $P(\rho)$ and $P(\rho, \mathbf{b}_j)$ as defined in (3.5) and (3.6). This will set the stage for asking whether we can distinguish different conjugates of H from each other or from the trivial subgroup.

Fix an irreducible representation ρ that acts on a vector space V. Then Fourier transforming the state

$$|H\rangle = \frac{1}{\sqrt{|H|}} \sum_{h \in H} |h\rangle$$

yields the coefficient

$$\widehat{H}(\rho) = \sqrt{\frac{d_\rho}{|H||G|}} \sum_{h \in H} \rho(h) = \sqrt{\frac{d_\rho |H|}{|G|}}\, \Pi_H \ ,$$

where $\Pi_H = (1/|H|) \sum_{h \in H} \rho(h)$ is a projection operator onto a subspace of V. The probability that we observe ρ is then the norm squared of this coefficient,

$$P(\rho) = \left\| \widehat{H}(\rho) \right\|^2 = \frac{d_\rho |H|}{|G|}\, \mathbf{rk}\, \Pi_H \ , \tag{3.7}$$

and, as stated above, this is the same for all conjugates H^g. The conditional probability that we observe the vector $\mathbf{b}_j$, given that we observe ρ, is then

$$P(\rho, \mathbf{b}_j) = a_j \frac{\left\| \Pi^\rho_{\mathbf{b}_j} |H\rangle \right\|^2}{P(\rho)} = a_j \frac{\left\| \widehat{H}(\rho) \mathbf{b}_j \right\|^2}{P(\rho)} = a_j \frac{\|\Pi_H \mathbf{b}_j\|^2}{\mathbf{rk}\, \Pi_H} \ . \tag{3.8}$$

In the case where H is the trivial subgroup, $\Pi_H = \mathbb{1}_{d_\rho}$ and $P(\rho, \mathbf{b}_j)$ is given by

$$P(\rho, \mathbf{b}_j) = \frac{a_j}{d_\rho} \ . \tag{3.9}$$

We call this the *natural distribution* on the frame $B = \{\mathbf{b}_j\}$. In the case that B is an orthonormal basis, $a_j = 1$ and $P(\rho, \mathbf{b}_j)$ is simply the uniform distribution on B.

This probability distribution over B changes for a conjugate H^g in the following way. The Fourier transform of $|Hg\rangle$ is

$$\widehat{Hg}(\rho) = \sqrt{\frac{d_\rho |H|}{|G|}}\, \Pi_H \rho(g) \ ,$$

and we have

$$P(\rho, \mathbf{b}_j) = a_j \frac{\|\Pi_H(g\mathbf{b}_j)\|^2}{\mathbf{rk}\, \Pi_H} \ , \tag{3.10}$$

where we write $g\mathbf{b}$ for $\rho(g)\mathbf{b}$.

Our goal is to understand, for each fixed $\mathbf{b}$, to what extent $P(\rho, \mathbf{b})$ varies with g, and so to what extent measurements of this type can distinguish the conjugates H^g from each other. Regarding this as a random variable over the choice of g, its expectation is easy to calculate: we have

$$\mathrm{Exp}_g \|\Pi_H(g\mathbf{b})\|^2 = \mathrm{Exp}_g \left\langle \mathbf{b}, \rho(g)^\dagger \Pi_H \rho(g) \mathbf{b} \right\rangle = \left\langle \mathbf{b}, \left(\mathrm{Exp}_g\, \rho(g)^\dagger \Pi_H \rho(g)\right) \mathbf{b} \right\rangle = \frac{\mathbf{rk}\, \Pi_H}{d_\rho} \ ,$$

where we used (2.1), $\|\mathbf{b}\|^2 = 1$, and the fact that the trace of a projection operator is its rank. Combining this with (3.10), the expected probability is simply the natural distribution (3.9),

$$\mathrm{Exp}_g\, P(\rho, \mathbf{b}_j) = \frac{a_j}{d_\rho} \ .$$

We wish to show that $\|\Pi_H(g\mathbf{b})\|^2$, and therefore $P(\rho, \mathbf{b}_j)$, is in fact very close to its expectation for most conjugates. In the next section, we present our primary technical contribution, which is a method for establishing concentration results for this random variable.

4. The variance of projection through a random involution. In this section we focus on the case where $H = \{1, m\}$ for an element m chosen uniformly at random from a fixed conjugacy class $\mathfrak{m}$ of involutions. (Observe that order is preserved under conjugation so that if m is an involution, then so are all elements of $\mathfrak{m}$.) Given an irreducible representation $\rho : G \to \mathsf{U}(V)$ and a vector $\mathbf{b} \in V$, we bound the variance, over the choice of $m \in \mathfrak{m}$, of the probability $P_m(\rho, \mathbf{b})$ that $\mathbf{b}$ is observed given that we observed ρ. Our key insight is that this variance depends on how the tensor product representation $\rho \otimes \rho^*$ decomposes into irreducible representations σ, and how the vector $\mathbf{b} \otimes \mathbf{b}^*$ projects into these constituent subspaces.

Recall that, if a representation ρ is reducible, it can be written as an orthogonal direct sum of irreducibles $\rho = \bigoplus_{\sigma \prec \rho} a_\sigma^\rho \sigma$, where a_σ^ρ is the multiplicity of σ. We let Π_σ^ρ denote the projection operator whose image is $a_\sigma^\rho \sigma$, that is, the span of all the irreducible subspaces isomorphic to σ.

LEMMA 4.1. *Let ρ be a representation of a group G acting on a space V and let $\mathbf{b} \in V$. Let m be an element chosen uniformly from a conjugacy class $\mathfrak{m}$ of involutions. If ρ is irreducible, then*

$$\mathrm{Exp}_{m\in\mathfrak{m}} \langle \mathbf{b}, m\mathbf{b}\rangle = \frac{\chi_\rho(\mathfrak{m})}{d_\rho} \|\mathbf{b}\|^2 \quad .$$

If ρ is reducible, then

$$\mathrm{Exp}_{m\in\mathfrak{m}} \langle \mathbf{b}, m\mathbf{b}\rangle = \sum_{\sigma\prec\rho} \frac{\chi_\sigma(\mathfrak{m})}{d_\sigma} \|\Pi_\sigma^\rho \mathbf{b}\|^2 \quad .$$

Proof. Let $\rho(\mathfrak{m})$ denote the average of ρ over the conjugacy class $\mathfrak{m}$. Using (2.2), we have

$$\mathrm{Exp}_m \langle \mathbf{b}, m\mathbf{b}\rangle = \langle \mathbf{b}, \rho(\mathfrak{m})\mathbf{b}\rangle = \frac{\chi_\rho(\mathfrak{m})}{d_\rho} \|\mathbf{b}\|^2 \quad .$$

Similarly, if ρ is reducible, by (2.3) we have

$$\mathrm{Exp}_m \langle \mathbf{b}, m\mathbf{b}\rangle = \langle \mathbf{b}, \rho(\mathfrak{m})\mathbf{b}\rangle = \sum_{\sigma\prec\rho} \frac{\chi_\sigma(\mathfrak{m})}{d_\sigma} \langle \mathbf{b}, \Pi_\sigma^\rho \mathbf{b}\rangle = \sum_{\sigma\prec\rho} \frac{\chi_\sigma(\mathfrak{m})}{d_\sigma} \|\Pi_\sigma^\rho \mathbf{b}\|^2 \quad . \qquad \square$$

Turning now to the second moment of $\langle \mathbf{b}, m\mathbf{b}\rangle$, we observe that

$$|\langle \mathbf{b}, m\mathbf{b}\rangle|^2 = \langle \mathbf{b}, m\mathbf{b}\rangle \langle \mathbf{b}, m\mathbf{b}\rangle^* = \langle \mathbf{b}\otimes\mathbf{b}^*, m\mathbf{b}\otimes m\mathbf{b}^*\rangle = \langle \mathbf{b}\otimes\mathbf{b}^*, m(\mathbf{b}\otimes\mathbf{b}^*)\rangle \ ,$$

where the action of m on the vector $\mathbf{b}\otimes\mathbf{b}^*$ is precisely given by the action of G in the representations $\rho\otimes\rho^*$. This will allow us to express the second moment of the inner product $\langle \mathbf{b}, m\mathbf{b}\rangle$ in terms of the projections of $\mathbf{b}\otimes\mathbf{b}^*$ into the irreducible constituents of the tensor product representation $\rho\otimes\rho^*$.

LEMMA 4.2. *Let ρ be a representation of a group G acting on a space V and let $\mathbf{b} \in V$. Let m be an element chosen uniformly at random from a conjugacy class $\mathfrak{m}$ of involutions. Then*

$$\mathrm{Exp}_{m\in\mathfrak{m}} |\langle \mathbf{b}, m\mathbf{b}\rangle|^2 = \sum_{\sigma\prec\rho\otimes\rho^*} \frac{\chi_\sigma(\mathfrak{m})}{d_\sigma} \left\| \Pi_\sigma^{\rho\otimes\rho^*} (\mathbf{b}\otimes\mathbf{b}^*) \right\|^2 \quad .$$

Proof. We write the second moment as a first moment over the product representation $\rho\otimes\rho^*$: as above, $|\langle \mathbf{b}, m\mathbf{b}\rangle|^2 = \langle \mathbf{b}\otimes\mathbf{b}^*, m(\mathbf{b}\otimes\mathbf{b}^*)\rangle$ so that

$$\mathrm{Exp}_m |\langle \mathbf{b}, m\mathbf{b}\rangle|^2 = \mathrm{Exp}_m \langle \mathbf{b}\otimes\mathbf{b}^*, m(\mathbf{b}\otimes\mathbf{b}^*)\rangle \ ,$$

and applying Lemma 4.1 completes the proof. $\square$

Now let $\Pi_m = \Pi_H$ denote the projection operator given by

$$\Pi_m \mathbf{v} = \frac{\mathbf{v} + m\mathbf{v}}{2} \quad .$$

For a given vector $\mathbf{b} \in B$, we will focus on the expectation and variance of $\|\Pi_m \mathbf{b}\|^2$. These are given by the following lemma.

LEMMA 4.3. *Let ρ be an irreducible representation acting on a space V and let $\mathbf{b} \in V$. Let m be an element chosen uniformly at random from a conjugacy class $\mathfrak{m}$ of involutions. Then*

$$\mathrm{Exp}_{m\in\mathfrak{m}} \|\Pi_m \mathbf{b}\|^2 = \frac{1}{2}\|\mathbf{b}\|^2 \left(1 + \frac{\chi_\rho(\mathfrak{m})}{d_\rho}\right) \quad , \tag{4.1}$$

$$\mathrm{Var}_{m\in\mathfrak{m}} \|\Pi_m \mathbf{b}\|^2 \le \frac{1}{4} \sum_{\sigma \prec \rho\otimes\rho^*} \frac{\chi_\sigma(\mathfrak{m})}{d_\sigma} \left\|\Pi_\sigma^{\rho\otimes\rho^*}(\mathbf{b}\otimes\mathbf{b}^*)\right\|^2 \quad . \tag{4.2}$$

Proof. For the expectation,

$$\begin{aligned}
\mathrm{Exp}_m \|\Pi_m \mathbf{b}\|^2 &= \mathrm{Exp}_m \langle \mathbf{b}, \Pi_m \mathbf{b}\rangle \\
&= \frac{1}{2}\mathrm{Exp}_m \left(\langle \mathbf{b}, \mathbf{b}\rangle + \langle \mathbf{b}, m\mathbf{b}\rangle\right) \\
&= \frac{1}{2}\|\mathbf{b}\|^2 \left(1 + \frac{\chi_\rho(\mathfrak{m})}{d_\rho}\right) \quad ,
\end{aligned}$$

where the last equality follows from Lemma 4.1.

For the variance, we first calculate the second moment,

$$\begin{aligned}
\mathrm{Exp}_m \|\Pi_m \mathbf{b}\|^4 &= \mathrm{Exp}_m |\langle \mathbf{b}, \Pi_m \mathbf{b}\rangle|^2 \\
&= \frac{1}{4}\mathrm{Exp}_m |\langle \mathbf{b}, \mathbf{b}\rangle + \langle \mathbf{b}, m\mathbf{b}\rangle|^2 \\
&= \frac{1}{4}\mathrm{Exp}_m \left(|\langle \mathbf{b}, \mathbf{b}\rangle|^2 + 2\,\mathrm{Re}\,\langle \mathbf{b}, \mathbf{b}\rangle\langle \mathbf{b}, m\mathbf{b}\rangle + |\langle \mathbf{b}, m\mathbf{b}\rangle|^2\right) \\
&= \frac{1}{4}\left(\|\mathbf{b}\|^4 + 2\|\mathbf{b}\|^4 \frac{\chi_\rho(\mathfrak{m})}{d_\rho} + \sum_{\sigma\prec\rho\otimes\rho^*} \frac{\chi_\sigma(\mathfrak{m})}{d_\sigma} \left\|\Pi_\sigma^{\rho\otimes\rho^*}(\mathbf{b}\otimes\mathbf{b}^*)\right\|^2\right) \quad ,
\end{aligned}$$

where in the last line we applied Lemmas 4.1 and 4.2 and the fact that any character evaluated at an involution is real. Then

$$\begin{aligned}
\mathrm{Var}_m \|\Pi_m \mathbf{b}\|^2 &= \mathrm{Exp}_m \|\Pi_m \mathbf{b}\|^4 - \left(\mathrm{Exp}_m \|\Pi_m \mathbf{b}\|^2\right)^2 \\
&= \frac{1}{4}\left[\sum_{\sigma\prec\rho\otimes\rho^*} \frac{\chi_\rho(\mathfrak{m})}{d_\rho} \left\|\Pi_\sigma^{\rho\otimes\rho^*}(\mathbf{b}\otimes\mathbf{b}^*)\right\|^2 - \|\mathbf{b}\|^4 \left(\frac{\chi_\rho(\mathfrak{m})}{d_\rho}\right)^2\right] \quad . \qquad (4.3)
\end{aligned}$$

Ignoring the second term, which is negative, gives the stated result. ∎

Finally, we point out that since

$$\mathrm{Exp}_m \|\Pi_m \mathbf{b}\|^2 = \|\mathbf{b}\|^2 \frac{\mathbf{rk}\,\Pi_m}{d_\rho} \quad ,$$

we have

$$\frac{\mathbf{rk}\,\Pi_m}{d_\rho} = \frac{1}{2}\left(1 + \frac{\chi_\rho(\mathfrak{m})}{d_\rho}\right) \quad , \tag{4.4}$$

a fact which we will use below.

5. The representation theory of the symmetric group. In this section we record the particular properties of S_n and its representation theory which we apply in the proofs of our main results. The irreducible representations of S_n are labeled by Young diagrams or, equivalently, by integer partitions of n,

$$\lambda = (\lambda_1, \ldots, \lambda_t) \ ,$$

where $\sum_i \lambda_i = n$ and $\lambda_i \geq \lambda_{i+1}$ for all i. The number of Young diagrams, equal to the number of conjugacy classes in S_n, is the partition number $p(n)$, which obeys

$$p(n) = (1 + o(1)) \frac{1}{4\sqrt{3} \cdot n} e^{\delta\sqrt{n}} < e^{\delta\sqrt{n}}, \quad \text{where} \quad \delta = \pi\sqrt{2/3} \ . \tag{5.1}$$

We identify each irreducible representation with its Young diagram λ, and denote its character χ_λ and its dimension d_λ. In particular, λ is the trivial or parity representation if λ is a single row (n) or a single column $(1, \ldots, 1)$, respectively. Given λ, its *conjugate* λ' is obtained by flipping λ about the diagonal: $\lambda' = (\lambda'_1, \ldots, \lambda'_{\lambda_1})$, where $\lambda'_j = |\{i \mid \lambda_i \geq j\}|$. In particular, $\lambda'_1 = t$. The representation λ' is the (tensor) product of λ with the parity representation.

The dimension of λ is given by the remarkable *hook length formula*:

$$d_\lambda = \frac{n!}{\prod_c \text{hook}(c)} \ ,$$

where this product runs over all cells of the Young diagram associated with λ and $\text{hook}(c)$ is the number of cells appearing in either the same column or row as c, excluding those that are above or to the left of c.

For example, the partition $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (6, 5, 3, 2)$ is associated with the diagram shown in Figure 5.1 below. The hook associated with the cell $(2, 2)$ in this diagram appears in Figure 5.2; it has length 6.

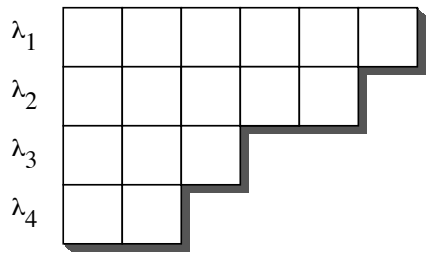

FIG. 5.1. *The Young diagram for* $\lambda = (6, 5, 3, 2)$.

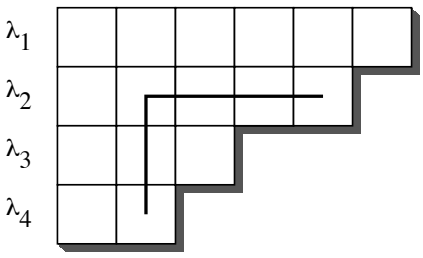

FIG. 5.2. *A hook of length* 6.

The symmetric groups have the property that every representation λ possesses a basis in which its matrix elements are real, and so all its characters are real. However, in a given basis λ might be complex, so we will refer below to its complex conjugate, the representation λ^* (not to be confused with λ').

The study of the asymptotic properties of the representations of S_n typically focuses on the *Plancherel* distribution (see, e.g., Kerov's monograph [23]). For a general group G, this is the probability distribution obtained on $\widehat{G}$ by assigning ρ the probability density $d_\rho^2/|G|$. One advantage of this distribution is that the density at

ρ is proportional to its contribution, dimensionwise, to the group algebra $\mathbb{C}[G]$. Note that in the context of the HSP, the Plancherel distribution is exactly the one obtained by performing weak Fourier sampling on the trivial hidden subgroup.

In the symmetric groups a fair amount is known about representations chosen according to the Plancherel distribution. In particular, Vershik and Kerov [36] have given the following result, showing that with high probability they have dimension equal to $e^{\Theta(\sqrt{n})}\sqrt{n!}$.

THEOREM 5.1 (see [36]). *Let λ be chosen from $\widehat{S_n}$ according to the Plancherel distribution. Then there exist positive constants c_1 and c_2 for which*

$$\lim_{n\to\infty} \Pr\left[e^{-c_1\sqrt{n}}\sqrt{n!} \le d_\lambda \le e^{-c_2\sqrt{n}}\sqrt{n!}\right] = 1 \ .$$

Vershik and Kerov have also obtained estimates for the *maximum* dimension of a representation in $\widehat{S_n}$.

THEOREM 5.2 (see [36]). *There exist positive constants $\check{c}$ and $\hat{c}$ such that for all $n \ge 1$,*

$$e^{-\check{c}\sqrt{n}}\sqrt{n!} \le \max_{\lambda\in\widehat{S_n}} d_\lambda \le e^{-\hat{c}\sqrt{n}}\sqrt{n!} \ .$$

Along with these estimates, we will use the following (one-sided) large-deviation versions of Theorem 5.1.

LEMMA 5.3. *Let λ be chosen according to the Plancherel distribution on $\widehat{S_n}$.*

1. *Let $\delta = \pi\sqrt{2/3}$ as in (5.1). Then for sufficiently large n,*

$$\Pr\left[d_\lambda \le e^{-\delta\sqrt{n}}\sqrt{n!}\right] < e^{-\delta\sqrt{n}} \ .$$

2. *Let $0 < c < 1/2$. Then there is a constant $\gamma > 0$ such that*

$$\Pr[d_\lambda \le n^{cn}] < n^{-\gamma n} \ .$$

Proof. For the first bound, setting $d = e^{-\delta\sqrt{n}}\sqrt{n!}$ and using (5.1), we have

$$\sum_{\lambda: d_\lambda \le d} \frac{d_\lambda^2}{n!} \le p(n)\frac{d^2}{n!} < e^{-\delta\sqrt{n}} \ .$$

For the second bound, recalling Stirling's approximation $n! > n^n e^{-n}$, we have

$$\sum_{\lambda: d_\lambda \le n^{cn}} \frac{d_\lambda^2}{n!} \le \frac{p(n)n^{2cn}}{n!} = n^{-(1-2c)n}e^{O(n)} \ ,$$

and setting $\gamma < 1 - 2c$ completes the proof. ∎

Finally, we will also apply Roichman's estimates [31] for the characters of the symmetric group.

DEFINITION 5.4. *For a permutation $\pi \in S_n$, define the* support *of π, denoted* $\mathrm{supp}(\pi)$*, to be the cardinality of the set $\{k \in [n] \mid \pi(k) \ne k\}$.*

THEOREM 5.5 (see [31]). *There exist constants $b > 0$ and $0 < q < 1$ so that for $n > 4$, for every conjugacy class C of S_n, and for every irreducible representation λ of S_n,*

$$\left|\frac{\chi_\lambda(C)}{d_\lambda}\right| \le \left(\max\left(q, \frac{\lambda_1}{n}, \frac{\lambda_1'}{n}\right)\right)^{b\cdot\mathrm{supp}(C)} \ ,$$

where $\mathrm{supp}(C) = \mathrm{supp}(\pi)$ *for any* $\pi \in C$.

In our application, we take n to be even and consider involutions m in the conjugacy class of elements consisting of $n/2$ disjoint transpositions, $\mathfrak{m} = \mathfrak{m}_n = \{\sigma\,((12)(34)\cdots(n-1\ \ n))\,\sigma^{-1} \mid \sigma \in S_n\}$. Note that each $m \in \mathfrak{m}_n$ is associated with one of the $(n-1)!! = (n-1)(n-3)(n-5)\cdots 1$ perfect matchings of n things, and that $\mathrm{supp}(m) = n$.

6. Strong Fourier sampling over S_n. We consider the hidden subgroup $H = \{1, m\}$, where m is chosen uniformly from $\mathfrak{m} = \mathfrak{m}_n \subset S_n$, the conjugacy class

$$\{\pi^{-1}((1\ 2)(3\ 4)\cdots(n-1\ n))\pi \mid \pi \in S_n\}\ ;$$

we assume throughout that n is even. We start by performing weak sampling, i.e., measuring the name of an irreducible representation λ; the resulting probability distribution on $\widehat{S_n}$ is the same for all $m \in M_n$, and Hallgren, Russell, and Ta-Shma [15] established that this probability distribution on λ is exponentially close to the Plancherel distribution in total variation. We continue on to strong sampling, by allowing the algorithm designer to choose an arbitrary POVM with a frame $B = \{\mathbf{b}_j\}$ and weights $\{a_j\}$ obeying the completeness condition (3.3). We will show that with high probability (over m and λ), the conditional distribution induced on the vectors B is exponentially close to the natural distribution (3.9) on B. It will follow by the triangle inequality that it requires an exponential number of experiments of this type to distinguish two involutions from each other or, in fact, to distinguish H from the trivial subgroup.

For simplicity, and to illustrate our techniques, we first prove this for a von Neumann measurement, i.e., where B is an orthonormal basis for λ. In this case, we show that the probability distribution on B is exponentially close to the uniform distribution.

6.1. Von Neumann measurements.

THEOREM 6.1. *Let B be an orthonormal basis for an irreducible representation λ. Given the hidden subgroup $H = \{1, m\}$, where m is chosen uniformly at random from $\mathfrak{m}$, let $P_m(\mathbf{b}) = P_m(\lambda, \mathbf{b})$ be the probability that we observe the vector $\mathbf{b} \in B$ conditioned on having observed the representation name λ, and let $U(\mathbf{b}) = U(\lambda, \mathbf{b})$ be the uniform distribution on B. Then there is a constant $\beta > 0$ such that for sufficiently large n, with probability at least $1 - e^{-\beta n}$ in m and λ, we have*

$$\|P_m - U\|_1 < e^{-\beta n}\ .$$

Proof. First, recall from (3.8) in section 3 that the conditional distribution on B is given by (since $a_j = 1$)

$$P_m(\mathbf{b}) = P_m(\lambda, \mathbf{b}) = \frac{\|\Pi_m \mathbf{b}\|^2}{\mathbf{rk}\,\Pi_m}\ . \tag{6.1}$$

Our strategy will be to bound $\mathrm{Var}_m \|\Pi_m \mathbf{b}\|^2$ using Lemma 4.3 and apply Chebyshev's inequality to conclude that $\|\Pi_m \mathbf{b}\|^2$ is almost certainly close to its expectation (4.1). Recall, however, that our bounds on the variance of $\|\Pi_m \mathbf{b}\|^2$ depend on the decomposition of $\lambda \otimes \lambda^*$ into irreducibles and, furthermore, on the projection of $\mathbf{b} \otimes \mathbf{b}^*$ into these irreducible subspaces. Matters are somewhat complicated by the fact that certain irreducibles μ appearing in $\lambda \otimes \lambda^*$ may contribute more to the variance than others. Specifically, while Theorem 5.5 allows us to bound the contribution of those μ

with Young diagrams whose width μ_1 and height μ'_1 are much smaller than n, those which violate this condition could have large normalized characters $\chi_\mu(\mathfrak{m})/d_\mu$, and thus could conceivably contribute large terms to the sum (4.2).

Fortunately, we will see that the total fraction of the space $\lambda \otimes \lambda^*$, dimensionwise, consisting of such μ is small with overwhelming probability. Despite this, we cannot preclude the possibility that for a *specific* vector $\mathbf{b}$, the quantity $\mathrm{Var}\,\|\Pi_m \mathbf{b}\|^2$ is large, as $\mathbf{b}$ may project solely into spaces of the type described above. On the other hand, as these troublesome spaces amount to a small fraction of $\lambda \otimes \lambda^*$, only a few $\mathbf{b}$ can have this property, and this will suffice to control the distance in total variation from the uniform distribution.

Specifically, let $0 < c < 1/4$ be a constant, and let Λ_c denote the collection of Young diagrams μ with the property that either $\mu_1 \geq (1-c)n$ or $\mu'_1 \geq (1-c)n$. We have the following upper bounds on the cardinality of Λ_c and the dimension of any μ with $\mu \in \Lambda_c$.

LEMMA 6.2. *Let $p(n)$ denote the number of integer partitions of n. Then $|\Lambda_c| \leq 2cn \cdot p(cn)$, and $d_\mu < n^{cn}$ for any $\mu \in \Lambda_c$.*

Proof. For the first statement, note that removing the top row of a Young diagram μ with $\mu_1 \geq (1-c)n$ gives a Young diagram of size $n - \mu_1 \leq cn$. The number of these is at most $p(cn)$, and summing over all such μ_1 gives $cn \cdot p(cn)$. The case $\mu'_1 \geq (1-c)n$ is similar, and summing the two gives $|\Lambda_c| \leq 2cn \cdot p(cn)$.

Now let $\mu \in \Lambda_c$ with $\mu_1 \geq (1-c)n$. By the hook-length formula, since the ith cell from the right in the top row has $\mathrm{hook}(c) \geq i$, $d_\mu < n!/\mu_1! \leq n!/((1-c)n)! \leq n^{cn}$. The case $\mu'_1 \geq (1-c)n$ is similar. ∎

To introduce a bit more notation, given a constant d, let M_d denote the set of irreducibles λ such that $d_\lambda \leq n^{dn}$. Now Lemma 5.3, part 2 shows that if λ is drawn according to the Plancherel distribution, the probability that it falls into M_d for some $d < 1/2$ is $n^{-\Omega(n)}$. The following lemma shows that this is also true for the distribution $P(\rho)$ induced on $\widehat{S_n}$ by weak Fourier sampling the coset state $|H\rangle$.

LEMMA 6.3. *Let $d < 1/2$ be a constant and let λ be drawn according to the distribution $P(\cdot)$ of (3.7). Then there is a constant $\gamma = \gamma(d) > 0$ such that for sufficiently large n we have $\Pr_\lambda[d_\lambda \in \mathrm{M}_d] \leq n^{-\gamma n}$.*

Proof. As $|H| = 2$, $|G| = n!$, and $\mathbf{rk}\,\Pi_H \leq d_\rho$, we have

$$P(\rho) = \frac{d_\rho |H|}{|G|}\,\mathbf{rk}\,\Pi_H \leq \frac{2d_\rho^2}{n!}\ .$$

Thus $P(\cdot)$ is at most twice the Plancherel measure, and applying Lemma 5.3, part 2 completes the proof. ∎

Now, for a representation μ with $\mu \notin \Lambda_c$, Theorem 5.5 implies that

$$\left|\frac{\chi_\mu(\mathfrak{m})}{d_\mu}\right| \leq \big(\max(q, 1-c)\big)^{bn} \leq e^{-\alpha n} \tag{6.2}$$

for a constant $\alpha > 0$. Thus the contribution of such an irreducible to the variance estimate of Lemma 4.3 is exponentially small. In addition, note that Lemma 6.2 implies that $\Lambda_c \subset \mathrm{M}_d$ so long as $d > c$; we shall in fact assume that $c < 1/4 < d$ (and, moreover, that $2c < d$) so that conditioning on $\lambda \notin \mathrm{M}_d$, (4.4) and (6.2) imply that

$$\frac{d_\lambda}{2}\left(1 - e^{-\alpha n}\right) \leq \mathbf{rk}\,\Pi_m \leq \frac{d_\lambda}{2}\left(1 + e^{-\alpha n}\right)\ . \tag{6.3}$$

We turn now to the problem of bounding the multiplicities with which representations $\mu \in \Lambda_c$ can appear in $\lambda \otimes \lambda^*$. While no explicit decomposition is known for $\lambda \otimes \lambda^*$, the *endomorphism representations* of S_n, we record a coarse bound below which will suffice for our purposes. Recall that a character of $\lambda \otimes \lambda^*$ is χ_λ^2 as the characters of S_n are real. The multiplicity of the representation μ in $\lambda \otimes \lambda^*$ is $\langle \chi_\mu, \chi_\lambda^2 \rangle_G$. However, this is equal to $\langle \chi_\mu \chi_\lambda, \chi_\lambda \rangle_G$, the multiplicity of λ in $\mu \otimes \lambda$. Counting dimensions, this is clearly no more than $\dim(\mu \otimes \lambda)/\dim \lambda = d_\mu$. Hence the multiplicity of μ in $\lambda \otimes \lambda^*$ is bounded by

$$\langle \chi_\mu, \chi_\lambda^2 \rangle_G \leq d_\mu \ . \tag{6.4}$$

Let $L \subset \lambda \otimes \lambda^*$ be the subspace consisting of copies of representations μ with $\mu \in \Lambda_c$, and let Π_L be the projection operator onto this subspace. By Lemma 6.2, we have

$$\dim L \leq \sum_{\mu \in \Lambda_c} d_\mu^2 \leq 2cn \cdot p(cn) \cdot n^{2cn} = n^{2cn} e^{O(\sqrt{n})} \ .$$

Moreover, as B is an orthonormal basis for λ, the vectors $\{\mathbf{b} \otimes \mathbf{b}^* \mid \mathbf{b} \in B\}$ are mutually orthogonal in $\lambda \otimes \lambda^*$. Therefore,

$$\sum_{\mathbf{b} \in B} \|\Pi_L(\mathbf{b} \otimes \mathbf{b}^*)\|^2 \leq \dim L \ . \tag{6.5}$$

Applying the general bound provided by Lemma 4.3 on the variance of $\|\Pi_m \mathbf{b}\|^2$ with the estimates (6.5) and (6.2) above, assuming pessimistically that $\chi_\mu(M)/d_\mu = 1$ for all $\mu \in \Lambda_c$, and assuming that $\lambda \notin M_d$ so that $|B| = d_\lambda > n^{dn}$ yields

(6.6)

$$\begin{aligned}
\frac{1}{d_\lambda} \sum_{\mathbf{b}} \mathrm{Var}_m \|\Pi_m \mathbf{b}\|^2 &\leq \frac{1}{4d_\lambda} \left[\sum_{\mathbf{b}} \sum_{\mu \in \Lambda_c} \left\| \Pi_\mu^\lambda (\mathbf{b} \otimes \mathbf{b}^*) \right\|^2 \right. \\
&\qquad \left. + \sum_{\mathbf{b}} \sum_{\mu \notin \Lambda_c} \frac{\chi_\mu(M)}{d_\mu} \left\| \Pi_\mu^\lambda (\mathbf{b} \otimes \mathbf{b}^*) \right\|^2 \right] \\
&\leq \frac{1}{4d_\lambda} \left[n^{2cn} e^{O(\sqrt{n})} + e^{-\alpha n} d_\lambda \right] \leq \frac{1}{4} \left(n^{(2c-d)n} e^{O(\sqrt{n})} + e^{-\alpha n} \right) \\
&\leq \frac{e^{-\alpha n}}{2} \ ,
\end{aligned}$$

for sufficiently large n.

We return to our goal of bounding $\|P(\lambda, \cdot) - U(\lambda, \cdot)\|_1$ for a typical λ. (We note that the following part of the proof is considerably simplified from the conference version of this paper and is similar to the argument in the multiregister case appearing in [16].) First, note that for $1/2 > d > 1/4 > c$ and sufficiently large n,

$$\begin{aligned}
\mathrm{Exp}_\lambda \mathrm{Exp}_m \|P_m(\lambda, \cdot) - U(\lambda, \cdot)\|_1^2 &\leq 4 \Pr[\lambda \in \mathrm{M}_d] + \max_{\lambda \notin \mathrm{M}_d} \mathrm{Exp}_m \|P_m(\lambda, \cdot) - U(\lambda, \cdot)\|_1^2 \\
&= 4n^{-\gamma n} + \max_{\lambda \notin \mathrm{M}_d} \mathrm{Exp}_m \left(\sum_{\mathbf{b}} \left| \frac{\|\Pi_m \mathbf{b}\|^2}{\mathbf{rk}\, \Pi_m} - \frac{1}{d_\lambda} \right| \right)^2 \\
&= 4n^{-\gamma n} + \max_{\lambda \notin \mathrm{M}_d} \frac{1}{(\mathbf{rk}\, \Pi_m)^2} \mathrm{Exp}_m \left(\sum_{\mathbf{b}} \left| \|\Pi_m \mathbf{b}\|^2 - \frac{\mathbf{rk}\, \Pi_m}{d_\lambda} \right| \right)^2 \qquad (6.7)
\end{aligned}$$

$$\leq 4n^{-\gamma n} + \max_{\lambda \notin \mathrm{M}_d} \frac{d_\lambda}{(\mathbf{rk}\,\Pi_m)^2} \operatorname{Exp}_m \sum_{\mathbf{b}} \left(\|\Pi_m \mathbf{b}\|^2 - \frac{\mathbf{rk}\,\Pi_m}{d_\lambda} \right)^2 \tag{6.8}$$

$$\leq 4n^{-\gamma n} + \max_{\lambda \notin \mathrm{M}_d} \frac{4}{(1-e^{-\alpha n})^2} \left[\frac{1}{d_\lambda} \sum_{\mathbf{b}} \operatorname{Exp}_m \left(\|\Pi_m \mathbf{b}\|^2 - \frac{\mathbf{rk}\,\Pi_m}{d_\lambda} \right)^2 \right] \ , \tag{6.9}$$

where (6.8) follows from (6.7) by the Cauchy–Schwarz inequality and (6.9) follows from (6.8) by applying (6.3). Now observe that the bracketed expression is exactly that bounded by (6.6) above. Thus we have

$$\operatorname{Exp}_{\lambda,m} \|P_m(\lambda,\cdot) - U(\lambda,\cdot)\|_1^2 \leq 4n^{-\gamma n} + \frac{2e^{-\alpha n}}{(1-e^{-\alpha n})^2} \leq 3e^{-\alpha n}$$

for sufficiently large n. Finally, the assertion of the theorem follows by applying Markov's inequality and setting $\beta < \alpha/3$. □

6.2. Arbitrary POVMs. We now generalize the proof of Theorem 6.1 to the case where the algorithm designer is allowed to choose an arbitrary finite frame $B = \{\mathbf{b}\}$ of unit length vectors in λ, with a family of positive real weights $a_{\mathbf{b}}$ that satisfy the completeness condition

$$\sum_{\mathbf{b}} a_{\mathbf{b}} \, |\mathbf{b}\rangle \, \langle \mathbf{b}| = \mathbb{1} \ . \tag{6.10}$$

(Note that this is simply (3.3) where we have written $\mathbf{b}$ and $a_{\mathbf{b}}$ instead of $\mathbf{b}_j$ and a_j.)

THEOREM 6.4. *Let B be a frame with weights $\{a_{\mathbf{b}} \mid \mathbf{b} \in B\}$ satisfying the completeness condition* (6.10) *for an irreducible representation λ. Given the hidden subgroup $H = \{1, m\}$, where m is chosen uniformly at random from $\mathfrak{m}$, let $P_m(\mathbf{b}) = P_m(\lambda, \mathbf{b})$ be the probability that we observe the vector $\mathbf{b}$ conditioned on having observed the representation name λ, and let $N(\mathbf{b}) = N(\lambda, \mathbf{b})$ be the natural distribution* (3.9) *on B. Then there is a constant $\beta > 0$ such that for sufficiently large n, with probability at least $1 - e^{-\beta n}$ in m and λ, we have*

$$\|P_m - N\|_1 < e^{-\beta n} \ .$$

Proof. The proof of Theorem 6.1 goes through with a few modifications. Recall from (3.8) in section 3 that the conditional distribution on B is given by

$$P_m(\mathbf{b}) = P_m(\lambda, \mathbf{b}) = a_{\mathbf{b}} \frac{\|\Pi_m \mathbf{b}\|^2}{\mathbf{rk}\,\Pi_m} \ ,$$

and the natural distribution (3.9) is given by $N(\mathbf{b}) = a_{\mathbf{b}}/d_\lambda$.

First, let us change some semantics: given a subset $A \subseteq B$, we let $|A|$ denote the weighted size of A,

$$|A| = \sum_{\mathbf{b} \in A} a_{\mathbf{b}} \ .$$

With this definition, the total probability that falls in A under the natural distribution is $N(A) = |A|/d_\lambda$. With Λ_c and M_d defined as before, Lemmas 6.2 and 6.3 still apply. As in the development leading to (6.9), we find that $\operatorname{Exp}_\lambda \operatorname{Exp}_m \|P_m(\lambda,\cdot) - N(\lambda,\cdot)\|_1^2$ is no more than

$$4n^{-\gamma n} + \max_{\lambda \notin \mathrm{M}_d} \frac{1}{(\mathbf{rk}\,\Pi_m)^2} \operatorname{Exp}_m \left(\sum_{\mathbf{b}} a_{\mathbf{b}} \left| \|\Pi_m \mathbf{b}\|^2 - \frac{\mathbf{rk}\,\Pi_m}{d_\lambda} \right| \right)^2 \tag{6.11}$$

$$\leq 4n^{-\gamma n} + \max_{\lambda \notin \mathrm{M}_d} \frac{d_\lambda}{(\mathbf{rk}\,\Pi_m)^2} \operatorname{Exp}_m \sum_{\mathbf{b}} a_{\mathbf{b}} \left(\|\Pi_m \mathbf{b}\|^2 - \frac{\mathbf{rk}\,\Pi_m}{d_\lambda} \right)^2 \tag{6.12}$$

$$\leq 4n^{-\gamma n} + \max_{\lambda \notin \mathrm{M}_d} \frac{4}{(1-e^{-\alpha n})^2} \left[\frac{1}{d_\lambda} \sum_{\mathbf{b}} a_{\mathbf{b}} \operatorname{Exp}_m \left(\|\Pi_m \mathbf{b}\|^2 - \frac{\mathbf{rk}\,\Pi_m}{d_\lambda} \right)^2 \right] \ , \tag{6.13}$$

where (6.12) follows from (6.11) by applying the Cauchy–Schwarz inequality in the following way: for any function $f(\mathbf{b})$ we have

$$\left(\sum_{\mathbf{b}} a_{\mathbf{b}} |f(\mathbf{b})| \right)^2 \leq \left(\sum_{\mathbf{b}} a_{\mathbf{b}} \right) \left(\sum_{\mathbf{b}} a_{\mathbf{b}} \, |f(\mathbf{b})|^2 \right) = d_\lambda \sum_{\mathbf{b}} a_{\mathbf{b}} \, |f(\mathbf{b})|^2 \ .$$

As before, let $L \subset \lambda \otimes \lambda^*$ be the subspace consisting of copies of representations $\mu \in \Lambda_c$. In order to control the variance appearing in the bracketed expression of (6.13), we require an analogue of (6.5) for frames, proved below.

LEMMA 6.5. *Let L be a subspace of $\lambda \otimes \lambda^*$, and let Π_L be the projection operator onto L. Then*

$$\sum_{\mathbf{b}} a_{\mathbf{b}} \, \|\Pi_L(\mathbf{b} \otimes \mathbf{b}^*)\|^2 \leq \dim L \ . \tag{6.14}$$

Proof. First note that a vector $\mathbf{e} \in \lambda \otimes \lambda^*$ has entries $\mathbf{e}_{j,k}$ for $1 \leq j,k \leq d_\lambda$. There is a unique linear operator E on λ whose matrix entries are $E_{j,k} = \mathbf{e}_{j,k}$, and the inner product $\langle \mathbf{b} \otimes \mathbf{b}^*, \mathbf{e} \rangle$ in $\lambda \otimes \lambda^*$ can then be written as the bilinear form $\langle \mathbf{b}, E\mathbf{b} \rangle$ in λ. The Frobenius norm of E is $\|E\|^2 = \mathbf{tr}\, E^\dagger E = \|\mathbf{e}\|^2$.

Now let $\{\mathbf{e}_i\}$ be an orthonormal basis for L and let E_i be the operator corresponding to $\mathbf{e}_i$. Then

$$\begin{aligned} \sum_{\mathbf{b}} a_{\mathbf{b}} \, |\langle \mathbf{b} \otimes \mathbf{b}^*, \mathbf{e}_i \rangle|^2 &= \sum_{\mathbf{b}} a_{\mathbf{b}} \, |\langle \mathbf{b}, E_i \mathbf{b} \rangle|^2 \leq \sum_{\mathbf{b}} a_{\mathbf{b}} \, \|\mathbf{b}\|^2 \, \|E_i \mathbf{b}\|^2 = \sum_{\mathbf{b}} a_{\mathbf{b}} \, \|E_i \mathbf{b}\|^2 \\ &= \sum_{\mathbf{b}} a_{\mathbf{b}} \, \mathbf{tr} \left(E_i^\dagger \, |\mathbf{b}\rangle \, \langle \mathbf{b}| \, E_i \right) = \mathbf{tr} \left[E_i^\dagger \left(\sum_{\mathbf{b}} a_{\mathbf{b}} \, |\mathbf{b}\rangle \, \langle \mathbf{b}| \right) E_i \right] \\ &= \mathbf{tr}\, E_i^\dagger E_i = \|\mathbf{e}_i\|^2 = 1 \ , \end{aligned}$$

where we used the Cauchy–Schwarz inequality in the first line and completeness in the second line. Summing over the $\dim L$ basis vectors $\mathbf{e}_i$ then gives (6.14). ∎

Applying this lemma, we control $1/d_\lambda \cdot \sum_{\mathbf{b}} a_{\mathbf{b}} \operatorname{Var}_m \|\pi_m \mathbf{b}\|^2$ just as in (6.6). Substituting this bound for the bracketed expression of (6.13) and selecting $\beta < \alpha/3$ (as above) completes the proof. ∎

7. Structured involutions and the case of graph isomorphism. The preceding development focuses on the case where the hidden subgroup is distributed uniformly among the conjugates of the subgroup $H = \{1, m\}$. As such, this shows that the canonical reduction of GRAPH AUTOMORPHISM (the problem of determining whether a given graph has a nontrivial automorphism) to the HSP does not give rise to an efficient quantum algorithm via Fourier sampling.

However, the canonical reduction of Graph Isomorphism to the HSP induces a more structured set of involutions. As referred to in the introduction, fixing two rigid graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, each with n vertices, the automorphism group of their disjoint union $(V_1 \cup V_2,\ E_1 \cup E_2)$ is nontrivial exactly when they are isomorphic, in which case it is generated by an involution m with full support such that $m(V_1) = V_2$ and $m(V_2) = V_1$. Identifying V_1 and V_2 with the sets $\{1, \ldots, n\}$ and $\{n+1, \ldots, 2n\}$, respectively, and letting s denote the involution $(1\ n+1)(2\ n+2)\ldots(n\ 2n)$, the standard reduction to the HSP in S_{2n} then results in a hidden subgroup $H = \{1, m\}$, where m is a conjugate involution $a^{-1}sa$. However, rather than a being drawn from all of S_{2n}, it is an element of the Young subgroup $S_{n,n}$ which fixes V_1 and V_2:

$$S_{n,n} = \big\{\pi \in S_{2n} \mid \pi(\{1, \ldots, n\}) = \{1, \ldots, n\}\big\} \cong S_n \times S_n \ .$$

In other words, rather than considering all conjugates of m in S_{2n}, it suffices just to consider conjugates in $S_{n,n}$. A priori, it seems that this smaller set of possible hidden subgroups might be easier to identify. Moreover, let K be the subgroup generated by $S_{n,n}$ and s: this is the wreath product $S_n \wr \mathbb{Z}_2$, which can also be written as a semidirect product $K = (S_n \times S_n) \rtimes \mathbb{Z}_2$. Then each such H is contained in K, and it seems that it might be more intelligent to Fourier sample over K rather than over all of S_{2n}.

However, we can show that nothing is gained by this approach. First, note that the involutions described above form the (K-)conjugacy class

$$\big\{\big((\alpha, \alpha^{-1}), 1\big) \in (S_n \times S_n) \rtimes \mathbb{Z}_2 \mid \alpha \in S_n\big\} \ .$$

We remark that the development of section 3 is unchanged and that the optimal measurement to find a hidden conjugate again consists of strong Fourier sampling.

Now note that Fourier sampling over S_{2n} and over K is equivalent for the following reason: suppose we are trying to distinguish a set of hidden subgroups $H_i \subset G$, all of which are contained in a subgroup $K \subset G$. Let T be a set of representatives for the cosets of K. Then a random left coset of H_i in G is the product of a random left coset of H_i in K with a random element of T. Thus the mixed state describing a uniformly random coset of H_i in G can be written as the tensor product of the corresponding coset state over K with the completely mixed state over T. Since this completely mixed state (whose density matrix is the identity) contains no information, nothing is gained, or lost, by sampling over all of G rather than over K.

To proceed, we can determine K's irreducible representations and their characters, using the machinery of *induced representations* [33] as follows. For two irreducible representations ρ and σ of S_n, let $\rho \boxtimes \sigma$ denote their tensor product as a representation of $S_{n,n} \cong S_n \times S_n$. We consider the induced representation $\tau_{\{\rho,\sigma\}} = \mathrm{Ind}_{S_{n,n}}^{K}(\rho \boxtimes \sigma)$ and denote its character $\chi_{\{\rho,\sigma\}}$. It is easy to see that

$$\chi_{\{\rho,\sigma\}}\big(((\alpha,\beta),t)\big) = \begin{cases} \chi_\rho(\alpha)\chi_\sigma(\beta) + \chi_\sigma(\alpha)\chi_\rho(\beta) & \text{if } t = 0 \ , \\ 0 & \text{if } t = 1 \ ; \end{cases}$$

as the notation suggests, this depends only on the multiset $\{\rho, \sigma\}$. An easy computation shows that $\langle \chi_{\{\rho,\sigma\}}, \chi_{\{\rho,\sigma\}} \rangle = 1 + \delta_{\rho,\sigma}$. Thus, if $\rho \ncong \sigma$, then $\tau_{\{\rho,\sigma\}}$ is irreducible of dimension $2d_\rho d_\sigma$. On the other hand, if $\rho \cong \sigma$, then it decomposes into two irreducible representations of dimension d_ρ^2,

$$\tau_{\{\rho,\rho\}} \cong \tau_{\{\rho,\rho\},\mathbb{1}} \oplus \tau_{\{\rho,\rho\},\pi} \ , \tag{7.1}$$

where $\mathbb{1}$ and π are the trivial and sign representations, respectively, of $\mathbb{Z}_2$. Each of these irreducible representations acts on $V_\rho \otimes V_\rho$, the vector space supporting the action of $\rho \boxtimes \rho$. Both realize the element $((\alpha,\beta),0)$ as the linear map $\rho(\alpha) \otimes \rho(\beta)$, while $\tau_{\{\rho,\rho\},\mathbb{1}}$ and $\tau_{\{\rho,\rho\},\pi}$ realize the element $((1,1),1)$ as the maps which send $\mathbf{u} \otimes \mathbf{v}$ to $\mathbf{v} \otimes \mathbf{u}$ and $-\mathbf{v} \otimes \mathbf{u}$, respectively. The characters of these representations are

$$\begin{aligned} \chi_{\{\rho,\rho\},\mathbb{1}}\big(((\alpha,\beta),t)\big) &= \begin{cases} \chi_\rho(\alpha)\chi_\rho(\beta) & \text{if } t=0 \ , \\ \chi_\rho(\alpha\beta) & \text{if } t=1 \ , \end{cases} \\ \chi_{\{\rho,\rho\},\pi}\big(((\alpha,\beta),t)\big) &= \begin{cases} \chi_\rho(\alpha)\chi_\rho(\beta) & \text{if } t=0 \ , \\ -\chi_\rho(\alpha\beta) & \text{if } t=1 \ . \end{cases} \end{aligned} \tag{7.2}$$

In particular, since m is of the form $((\alpha,\alpha^{-1}),1)$, we have the normalized characters

$$\frac{\chi_{\{\rho,\rho\},\mathbb{1}}(m)}{d_{\{\rho,\rho\},\mathbb{1}}} = \frac{1}{d_\rho}, \quad \frac{\chi_{\{\rho,\rho\},\pi}(m)}{d_{\{\rho,\rho\},\pi}} = -\frac{1}{d_\rho} \ , \tag{7.3}$$

and $\chi_{\{\rho,\sigma\}}(m) = 0$ for all $\rho \ncong \sigma$.

Given that the normalized characters (7.3) are very small (indeed, $n^{-\Omega(n)}$) for all ρ whose Young diagram is outside Λ_c, the analysis of section 6 can be undertaken mutatis mutandis and easily implies that an exponential number of strong Fourier sampling experiments would have to be performed to distinguish the isomorphic and nonisomorphic cases. We note that a similar result has been obtained by Childs and Wojcan [6], who treat Graph Isomorphism as a hidden shift problem on S_n.

We remark that the above description (7.1), (7.2) of the irreducible representations and characters of groups of the form $G \wr \mathbb{Z}_2$ works for arbitrary G. In particular, the normalized characters of the involutions that "swap" the two copies of G are either 0 or $\pm 1/d_\rho$ for some $\rho \in \widehat{G}$. It follows that strong Fourier sampling fails to find such involutions in $G \wr \mathbb{Z}_2$ whenever a sufficient fraction of G's Plancherel measure lies on sufficiently high-dimensional representations.

Acknowledgments. We are grateful to Denis Thérien, McGill University, and Bellairs Research Institute for organizing a workshop at which this work began; to Dorit Aharonov, Daniel Rockmore, and Umesh Vazirani for helpful conversations; to Chris Lomont for pointing out several typos; and to Tracy Conrad and Sally Milius for their support and tolerance. C. M. also thanks Rosemary Moore for providing a larger perspective. Finally, we thank Gorjan Alagic for his comments on the structured involutions material.

REFERENCES

[1] G. Alagic, C. Moore, and A. Russell, *Quantum algorithms for Simon's problem over general groups*, in Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 2007, pp. 1217–1224.

[2] D. Bacon, A. Childs, and W. van Dam, *Optimal measurements for the dihedral hidden subgroup problem*, Chic. J. Theoret. Comput. Sci., (2006), article 2.

[3] D. Bacon, A. Childs, and W. van Dam, *From optimal measurement to efficient quantum algorithms for the hidden subgroup problem over semidirect product groups*, in Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science, 2005, pp. 469–478.

[4] R. Beals, *Quantum computation of Fourier transforms over symmetric groups*, in Proceedings of the 29th Annual ACM Symposium on Theory of Computing, 1997, pp. 48–53.

[5] E. Bernstein and U. Vazirani, *Quantum complexity theory* (preliminary abstract), in Proceedings of the 25th Annual ACM Symposium on Theory of Computing, 1993, pp. 11–20.
[6] A. Childs and P. Wojcan, *On the quantum hardness of solving isomorphism problems as nonabelian hidden shift problems*, Quantum Inf. Comput., 7 (2007), pp. 504–521.
[7] W. van Dam, S. Hallgren, and L. Ip, *Quantum algorithms for some hidden shift problems*, in Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 2003, pp. 489–498.
[8] M. Ettinger and P. Høyer, *On quantum algorithms for noncommutative hidden subgroups*, Adv. in Appl. Math., 25 (2000), pp. 239–251.
[9] M. Ettinger, P. Høyer, and E. Knill, *The quantum query complexity of the hidden subgroup problem is polynomial*, Inform. Process. Lett., 91 (2004), pp. 43–48.
[10] K. Friedl, G. Ivanyos, F. Magniez, M. Santha, and P. Sen, *Hidden translation and orbit coset in quantum computing*, in Proceedings of the 35th Annual ACM Symposium on Theory of Computing, 2003, pp. 1–9.
[11] W. Fulton and J. Harris, *Representation Theory: A First Course*, Grad. Texts in Math. 129, Springer-Verlag, New York, 1991.
[12] M. Grigni, L. J. Schulman, M. Vazirani, and U. Vazirani, *Quantum mechanical algorithms for the nonabelian hidden subgroup problem*, Combinatorica, 24 (2004), pp. 137–154.
[13] L. Hales and S. Hallgren, *Quantum Fourier sampling simplified*, in Proceedings of the 31st Annual ACM Symposium on Theory of Computing, 1999, pp. 330–338.
[14] L. Hales and S. Hallgren, *An improved quantum Fourier transform algorithm and applications*, in Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science, 2000, pp. 515–525.
[15] S. Hallgren, A. Russell, and A. Ta-Shma, *Normal subgroup reconstruction and quantum computation using group representations*, in Proceedings of the 32nd Annual ACM Symposium on Theory of Computing, 2000, pp. 627–635.
[16] S. Hallgren, C. Moore, M. Rötteler, A. Russell, and P. Sen, *Limitations of quantum coset states for graph isomorphism*, in Proceedings of the 38th Annual ACM Symposium on Theory of Computing, 2006, pp. 604–617.
[17] P. Høyer, *Efficient Quantum Transforms*, preprint, 1997; available online from http://arxiv.org/abs/quant-ph/9702028.
[18] Y. Inui and F. Le Gall, *An efficient algorithm for the hidden subgroup problem over a class of semi-direct product groups*, in Proceedings of EQIS, 2004.
[19] L. Ip, *Shor's Algorithm Is Optimal*, preprint, 2004.
[20] G. Ivanyos, F. Magniez, and M. Santha, *Efficient quantum algorithms for some instances of the non-abelian hidden subgroup problem*, Internat. J. Found. Comput. Sci., 14 (2003), pp. 723–740.
[21] R. Jozsa, *Quantum factoring, discrete logarithms and the hidden subgroup problem*, IEEE MultiMedia, 3 (1996), pp. 34-43.
[22] J. Kempe and A. Shalev, *The hidden subgroup problem and permutation group theory*, in Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 2005, pp. 1118–1125.
[23] S. V. Kerov, *Asymptotic Representation Theory of the Symmetric Group and Its Applications in Analysis*, Transl. Math. Monogr. 219, AMS, Providence, RI, 2003.
[24] G. Kuperberg, *A subexponential-time quantum algorithm for the dihedral hidden subgroup problem*, SIAM J. Comput., 35 (2005), pp. 170–188.
[25] C. Moore and A. Russell, *Explicit Multiregister Measurements for Hidden Subgroup Problems; or, Fourier Sampling Strikes Back*, preprint, 2005; available online from http://arxiv.org/abs/quant-ph/0504067.
[26] C. Moore and A. Russell, *For distinguishing conjugate hidden subgroups, the pretty good measurement is as good as it gets*, Quantum Inf. Commun., 7 (2007), pp. 752-765.
[27] C. Moore, D. Rockmore, and A. Russell, *Generic quantum Fourier transforms*, in Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 2004, pp. 771–780.
[28] C. Moore, D. Rockmore, A. Russell, and L. J. Schulman, *The power of basis selection in Fourier sampling: Hidden subgroup problems in affine groups*, in Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 2004, pp. 1106–1115.
[29] O. Regev, *Quantum computation and lattice problems*, in Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002, pp. 520–530.
[30] M. Rötteler and T. Beth, *Polynomial-Time Solution to the Hidden Subgroup Problem for a Class of Non-Abelian Groups*, preprint, 1998; available online from http://arxiv.org/abs/quant-ph/9812070.

[31] Y. Roichman, *Upper bound on the characters of the symmetric groups*, Invent. Math., 125 (1996), pp. 451–485.
[32] S. Roman, *Advanced Linear Algebra*, Grad. Texts in Math. 135, Springer-Verlag, New York, 1992.
[33] J.-P. Serre, *Linear Representations of Finite Groups*, Grad. Texts in Math. 42, Springer-Verlag, New York, 1977.
[34] P. W. Shor, *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer*, SIAM J. Comput., 26 (1997), pp. 1484–1509.
[35] D. R. Simon, *On the power of quantum computation*, SIAM J. Comput., 26 (1997), pp. 1474–1483.
[36] A. M. Vershik and S. V. Kerov, *Asymptotic behavior of the maximum and generic dimensions of irreducible representations of the symmetric group*, Funk. Anal. i Prolizhen, 19 (1985), pp. 25–36 (in Russian); Funct. Anal. Appl., 19 (1985), pp. 21–31 (in English).