Computing with Highly Mixed States

Andris Ambainis
School of Mathematics
Institute for Advanced Study
Princeton, NJ 08540
ambainis@ias.edu

Leonard J. Schulman
Department of Computer Science
California Institute of Technology
Pasadena, CA 91125
schulman@caltech.edu

Umesh Vazirani
Computer Science Division
University of California
Berkeley, CA 94720
vazirani@cs.berkeley.edu

Abstract

Device initialization is a difficult challenge in some proposed realizations of quantum computers, and as such, must be treated as a computational resource. The degree of initialization can be quantified by k, the number of clean qubits in the initial state of the register. In this paper we show that unless $m \in O(k + \log n)$, oblivious simulation of an ideal m-qubit quantum circuit by an n-qubit circuit with k clean qubits is impossible. Effectively, this indicates that there is no avoiding physical initialization of a quantity of qubits proportional to that required by the best ideal quantum circuit.

1 Introduction

The field of quantum computation is driven by the recent realization that there is a profound difference between quantum and classical mechanics in the resources (time, communication, interactive proof complexity, etc) required for computation. The peculiar nature of quantum mechanics also introduces new computational considerations that have no classical counterpart. In this paper, we study a computational resource — initialization — that is important in the setting of quantum computation, but which has little relevance in classical computation. Ideally, a quantum computation is a sequence of local unitary transformations applied to a register of qubits

which are initially in the state $|0^n\rangle$, followed by a measurement. However, initializing the state of the quantum register can be quite challenging. This is currently the limiting resource for certain implementations, such as NMR quantum computing. Initialization begins with a physical cooling (polarizing) stage, which results in each of the n qubits independently being in state $|0\rangle$ with probability $1/2 + \epsilon/2$ and in state $|1\rangle$ with probability $1/2 - \epsilon/2$. Current implementations in the laboratory all rely on a second phase of initialization which creates a pseudopure state which is a noisy version of $|0^n\rangle$. Unfortunately and unavoidably in this method, the signal to noise ratio decreases exponentially in n [13]. A different scheme for this second phase of initialization via algorithmic cooling was proposed in [18]. This scheme results in $m \approx \epsilon^2 n$ (almost) perfectly initialized qubits. However, currently achievable values of ϵ in liquid state NMR are around 10^{-5}, thus requiring n to be impractically large. Moreover, the bounds on m in [18] are tight to within lower order additive terms.

Is initialization to the state $|0^n\rangle$ really necessary for general quantum computation? This question was raised by Knill and Laflamme [10], who showed how to perform an interesting computation on a very weakly initialized quantum computer. The model they formalized is the k clean qubit model, which assumes that k of the qubits are perfectly initialized in the state $|0^k\rangle$, and the remaining $n - k$ are in the maximally mixed state (i.e., each qubit independently and uniformly distributed over $\{|0\rangle, |1\rangle\}$). Due to the algorithm [18] this model is essentially equivalent in NMR to a device in which the n qubits start out with bias $\epsilon = \sqrt{k/n}$. Knill and Laflamme adapted Kitaev's phase estimation technique to show how to use a 1-clean qubit device to estimate the trace of an n-qubit quantum circuit. No efficient classical algorithm is known for this task. This raised the possibility that preparation of strings of clean qubits is not a necessary prerequisite for running quantum algorithms.

This paper mostly dispels this possibility. With the exception of the Knill-Laflamme algorithm, all known quantum algorithms are designed for pure-state quantum computers. What we show is that any oblivious simulation of pure-state quantum algorithms on mixed-state quantum computers requires that the number of clean qubits k be almost as large as the number of qubits in the pure-state quantum algorithm (unless n is exponentially large). Thus we show that absent special purpose designs for mixed state algorithms, general purpose quantum computation in mixed state devices will hinge upon improvements in laboratory techniques for qubit cooling.

We begin on the positive side by showing that the one clean qubit model can be used to efficiently simulate any logarithmic depth classical circuit,

thus solving any problem in NC^1 efficiently. This is somewhat surprising because an easy argument shows that in the absence of clean qubits ($k = 0$), no computation at all is possible.

However, the primary question is whether the one clean qubit model (or k clean qubits for small k) be used to simulate a general quantum circuit on m qubits for m substantially larger than k. Our main result is negative — we show that an oblivious simulation of an m qubit quantum computation is possible in the k clean qubit model only if $m \leq (2k+\log n)(1+o(1))$. This is essentially, in fact, a geometric result: such a simulation requires a large collection (2^m) of distinguishable subspaces of dimension 2^{n-k} in $\mathbb{C}^{2^n}$ that can be permuted arbitrarily by unitary transformations. Using the representation theory of the symmetric group we show that the distinguishablity and permutability conditions cannot be simultaneously achieved unless m satisfies the bound above.

2 Preliminaries

If a physical system can be in N distinguishable (classical) states, then its (pure) quantum state is a linear superposition of these N states. Thus a two state system (called a qubit) is represented by a unit vector in a two dimensional complex Hilbert space:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \qquad \alpha, \beta \in \mathbb{C} \quad \text{and} \quad \alpha^2 + \beta^2 = 1$$

The state of an n qubit system is described by a unit vector in a 2^n dimensional Hilbert space:

$$|\psi\rangle = \sum_{x \in \{0,1\}^n} \alpha_x |x\rangle$$

Measuring the state of the system in the standard basis yields the outcome x with probability $|\alpha_x|^2$. Moreover, following the measurement, the new state of the system is $|x\rangle$. More generally, a measurement is specified by a set of projections $P_1, \ldots, P_m$ into orthogonal subspaces that span the underlying Hilbert space. The outcome of the measurement is j with probability $|P_j|\psi\rangle|^2$, and following the measurement, the new state of the system is $P_j|\psi\rangle$.

In actuality, there is generally some uncertainty about the state of the quantum system (due to entanglement with environment). This uncertainty is represented by a probability distribution over pure states. Such a system is referred to as a mixed state. There are, however, different distributions

over pure states that cannot be distinguished by any physical measurement. A complete description of a mixed state consisting of pure states $|\psi_i\rangle$ with probabilities p_i is its *density matrix,*

$$\sum_i p_i |\psi_i\rangle\langle\psi_i|.$$

Two mixed states are physically distinguishable if and only if they have different density matrices.

For example, the mixture of $|0\rangle$ with probability 1/2 and $|1\rangle$ with probability 1/2 has density matrix $\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$, as does the mixture of $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ with probability 1/2 and $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ with probability 1/2.

When two mixed states have distinct density matrices (say ρ_1 and ρ_2), it is natural to ask how well they can be distinguished by a measurement. The answer is that the least-error measurement is one that measures in the basis that diagonalizes $\rho_1 - \rho_2$. The success probability is proportional to the trace distance between ρ_1 and ρ_2, which is the sum of the absolute values of the eigenvalues of $\rho_1 - \rho_2$.

Diagonalizing the density matrix gives a mixture over orthogonal states. (Up to a unitary change of basis this can be regarded as a classical distribution). This is the mixture with minimum entropy consistent with the density matrix. The entropy of this distribution is invariant under reversible (unitary) quantum logic gates, and is referred to as the von Neumann entropy of the density matrix.

Quantum mechanics asserts that the evolution of a system is unitary. Each step amounts to multiplication of $|\psi\rangle$ by a certain unitary matrix U: $|\psi\rangle \to U|\psi\rangle$. In terms of density matrices, each computation step amounts to conjugation by U: $\rho \to U\rho U^{-1}$.

2.1 The basic model of a quantum computer

A quantum computer consists of the following:

- A quantum register which is initialized to some pure state $|\psi_0\rangle$, or more generally, to some density matrix ρ_0.
- The quantum algorithm is executed by performing a sequence of elementary gates on this quantum register. Each such gate is a unitary transformation that acts as the identity on all but two qubits. The

quantum computer is classically controlled, in two sense: first, the description of the sequence of elementary gates to be executed is produced using polynomial (classical) resources. Thus their complexity is not limited in any way by the number of clean qubits. Second, a classical controller applies these gates to the quantum register without becoming entangled with it.

- Generally it is important to allow for errors in the gates of the quantum computer. Since this paper is primarily concerned with a lower bound, we allow error-free gates.

- The output of the algorithm is obtained by measuring one or more qubits of the quantum register in the standard basis. Of course, the whole process may be repeated a polynomial number of times.

2.2 One qubit computation

In this paper, we consider the k-clean qubit model introduced by Knill and Laflamme [10]. The starting state of a 1-clean qubit computation is a uniform distribution over all basis states in which the first bit is a $|0\rangle$; this is given by the density matrix

$$\rho_{\text{start}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \otimes \underbrace{\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \otimes \ldots \otimes \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}}_{(n-1) \text{ times}}. \tag{1}$$

Equivalently, this is separable mixed state in which the first qubit is initialized to $|0\rangle$ while the rest are in states that are uniformly random ($|0\rangle$ with probability 1/2 and $|1\rangle$ with probability 1/2).

In the k-clean qubit model, k qubits are initialized to $|0\rangle$ while the rest are in uniformly random states.

If we have a state where every qubit is slightly biased towards $|0\rangle$ (as in the previous section), one can apply the purification procedure by Schulman and Vazirani [18] and transform it into a state where almost all the bias is concentrated on one qubit and the rest are almost completely random. Such a state is very close to the starting state for one slightly biased qubit computation. Therefore, any positive result about our model gives a corresponding positive result about NMR quantum computing.

We note that this is the weakest possible model of this type in which one can still do nontrivial computations. If all n qubits are completely random, no computation is possible. The start state is now

$$\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \otimes \ldots \otimes \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} = 2^{-n} I$$

where I is the $2^n \times 2^n$ identity matrix. Applying any computation U yields

$$U 2^{-n} I U^\dagger = 2^{-n} U U^\dagger = 2^{-n} I.$$

The final density matrix is therefore independent of the computation.

2.3 Subspace interpretation

A pure state is a vector $|\psi\rangle = \sum_{x_1 \ldots x_n} a_{x_1 \ldots x_n} |x_1 \ldots x_n\rangle$ in the Hilbert space $\mathbb{C}^{2^n}$. Similarly, the state of a k-clean qubit computer is described by a 2^{n-k} -dimensional subspace of $\mathbb{C}^{2^n}$. For instance, the initial state ρ_{start} in a one-clean qubit computer is described by the subspace V_{start} that is spanned by all basis states with first bit $= 0$. The identification is justified by the fact that the same ρ_{start} is defined by the uniform mixture of any orthonormal basis for the subspace. Applying U to this state transforms the density matrix to $U \rho_{\text{start}} U^\dagger$, and the subspace V_{start} to $U V_{\text{start}}$.

This correspondence will be extensively used in the proof of our lower bound on one-qubit computation (sections 6-8). For brevity, we will often refer to subspaces as the states of computation.

2.4 Representation theory

We now introduce the basic notions of representation theory needed to prove our main theorem. For more information on group representations, see [5, 17].

Representation. A representation ρ of a group G is a homomorphism ρ from G to the group of linear transformations $GL(V)$ of a vector space V. This means that, for any $g, h \in G$, $\rho(gh) = \rho(g)\rho(h)$. If the mapping ρ is clear from the context, we often call the space V itself a representation of G.

Irreducibility. We say that a subspace W is an *invariant* subspace of a representation ρ if $\rho(g)W \subseteq W$ for all $g \in G$. The zero subspace and the subspace V are always invariant. If no nonzero proper subspaces are invariant, the representation is said to be *irreducible*.

Isomorphism. Two representations $\rho : G \to GL(V)$ and $\rho' : G \to GL(W)$ are isomorphic if there is a bijective linear map $\varphi : V \to W$ such that $\varphi\rho(g) = \rho'(g)\varphi$ for any $g \in G$.

In this paper, we confine ourselves to finite groups and their representations over complex vector spaces. Under these conditions we have the following facts:

Unitarity. Every representation is conjugate to a unitary representation ρ, one in which every $\rho(g)$ is unitary. (So henceforth all representations will be assumed to be unitary).

Irreps. G has finitely many non-isomorphic irreducible representations (irreps), each of finite dimension. Every group has the trivial representation **1**of dimension 1, which maps all group elements to the number 1.

Schur's Lemma. If ρ and ρ' are two irreducible representations and φ is an isomorphism between them, then any other isomorphism φ' between ρ and ρ' is $c\varphi$ for some constant $c \in \mathbb{C}$.

Complete Reducibility Every representation V can be decomposed into irreps W_i possibly with multiplicity: $V = W_1 \oplus \ldots \oplus W_k$. This decomposition is unique when written as: $V = \sum W_i \otimes E_{j_i}$, where the W_i are non-isomorphic irreps, and $E_l = \oplus_{i=1}^{l} \mathbf{1}$. Here j_i is just the multiplicity with which the irrep W_i appears.

Irreducible representations of S_M. In this paper, we rely on the representation theory of the symmetric group S_M. The irreducible representations of S_M are in one-to-one correspondence with the partitions of n. A *partition* of M is a sequence $(\lambda_1, \ldots, \lambda_k)$ of positive integers, with $\lambda_1 \geq \ldots \geq \lambda_k$ for which $\sum \lambda_i = M$. It is customary to identify the partition $\lambda = (\lambda_1, \ldots, \lambda_k)$ with a diagram consisting of k rows of boxes, the ith row containing λ_i boxes. We will let λ stand for both the partition and the associated diagram. For example, the diagram corresponding to the partition $\lambda = (4, 4, 2, 1)$ is shown in figure 2.4.

The irreducible representation associated with λ is denoted ρ_λ. There is an explicit formula for the dimension of ρ_λ. This involves the notion of a *hook*: for a cell (i, j) of a Young tableau λ, the (i, j)-hook $h_{i,j}$ is the collection of all cells of λ which are beneath (i, j) (but in the same column) or to the right of (i, j) (but in the same row), including the

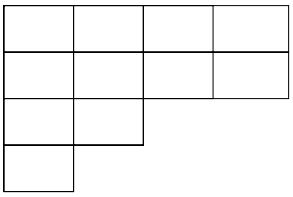

Figure 1: *The Young diagram of* $\lambda = (4, 4, 2, 1)$.

7	5	3	2
6	4	2	1
3	1		
1			

Figure 2: *The hook-lengths for (4, 4, 2, 1).*

cell (i, j). The *length* of the hook $\ell(h)$ is the number of cells appearing in the hook. With this notation, the dimension of ρ_λ may be expressed:

$$\dim \rho_\lambda = \frac{n!}{\prod_{i,j} \ell(h_{i,j})}, \tag{2}$$

this product being taken over all hooks h of λ. Figure 2.4 shows the hook lengths for the partition $\lambda = (4, 4, 3, 1)$. Formula (2) implies that the dimension of corresponding representation is

$$\frac{11!}{7 \cdot 5 \cdot 3 \cdot 2 \cdot 6 \cdot 4 \cdot 2 \cdot 3} = 1320.$$

Restriction. A representation ρ of a group G is also automatically a representation of any subgroup H. Note that even if a representation is irreducible over G, it may no longer be irreducible when restricted to H.

Restriction from S_M to S_{M-1}. In particular, we will be considering the restrictions of irreducible representations of S_M to S_{M-1} obtained by fixing an element $k \in \{1, \ldots, M\}$. Let λ be a partition of M and ρ_λ be the corresponding irreducible representation. Then, when we restrict

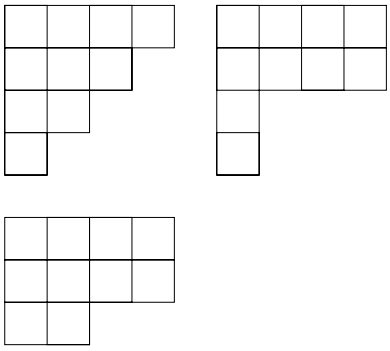

Figure 3: *The Young diagrams of representations of S_{10} contained in ρ_λ.*

to S_{M-1}, ρ_λ decomposes into irreducible representations of S_{M-1} in the following way:

$$\rho = \bigoplus_{\lambda^-} \rho_{\lambda^-}$$

where λ^- ranges over all shapes of size $M-1$ that can be obtained by deleting an "inside corner" from λ. (An inside corner is simply a point of the shape whose deletion leaves a legal shape.)

For example, the representation ρ_λ, $\lambda = (4,4,2,1)$ of S_{11} decomposes into 3 irreducible representations of S_{10}. The Young diagrams of these representations are shown in Fig. 2.4.

3 Positive result: NC^1

We show that there is a natural classical complexity class, NC^1, that can be computed in the one-clean-bit model using just three qubits. The fact that such nontrivial computations can be performed by one-clean-qubit devices underscores the difficulty of bounding the power of this class.

3.1 NC^1, branching programs and one-clean-qubit computation

NC^1 is the class of functions computable by a fanin-2 Boolean circuits of polynomial size and logarithmic depth. By Barrington's theorem [2], this class can be characterized by constant width permutation branching programs.

A branching program (BP) is a computing device that can be in one of finitely many states. It is specified by a sequence of instructions of the form (i, π_0, π_1) where i is an index for a variable x_i and π_0 and π_1 are permutations on the states. If $x_i = 0$, π_0 is applied. If $x_i = 1$, π_1 is applied. The length of a BP is the number of instructions in it.

For any given boolean inputs $x_1, \ldots, x_n$, running a BP induces the permutation $\pi_{x_n} \cdots \pi_{x_1}$ on the states. We say that a BP ω-computes a function $f(x_1, \ldots, x_n)$ if $\pi_{x_n} \cdots \pi_{x_1}$ is the identity when $f(x_1, \ldots, x_n) = 0$ and ω when $f = 1$.

Theorem 1 *[2] Let ω be a five-cycle. NC^1 is equal to the class of languages ω-recognized by a uniform family of width-5 polynomial length BPs.*

Proposition 1 *A three-qubit quantum computer initialized with one clean qubit can recognize every language in NC^1.*

Proof: Let $L \in NC^1$. Take the permutation branching program that ω-recognizes L where ω is the five-cycle $(000, 001, 010, 011, 100)$. We represent the five states of the permutation branching program by the states $|000\rangle$ through $|100\rangle$ of the three-qubit register.

Then, we simulate the permutation branching program by a sequence of elementary quantum operations. At the end, we measure the first qubit in the register. Two cases are possible:

1. $x \notin L$. Then, the branching program maps each of the states $|000\rangle$, $|001\rangle$, $|010\rangle$ and $|011\rangle$ to itself. This means that measuring the first qubit always gives 0.

2. $x \in L$. Then, the state $|011\rangle$ gets mapped to $|100\rangle$. Therefore, measuring the first qubit in the register yields a 1 with probability 1/4. □

This result can be refined in two respects. First, instead of getting 1 with probability 1/4 in the second case, we can get it with probability 1. To show that, we first notice that the proof of Barrington's theorem remains valid for eight-cycles instead five-cycles. Then, we take a BP that $(000, 100, 001, 101, 010, 110, 011, 111)$-recognizes L and apply the same argument. If $x \in L$, then the mixture of states $|000\rangle$, $|001\rangle$, $|010\rangle$, $|011\rangle$ is mapped to the mixture of $|100\rangle$, $|101\rangle$, $|110\rangle$, $|111\rangle$ and measuring the first qubit gives 1 with probability 1.

Second, our result was extended by Ablayev, Moore and Pollett [1] to show that the NC^1 simulation can be carried out on a computer with just one (clean) qubit.

3.2 QNC^1

It is illuminating to try to extend this simulation to QNC^1. QNC^1 is the class of languages recognized by quantum circuits of polynomial size and logarithmic size [12].

Notice that in Barrington's procedure for simulating NC^1, each wire in the NC^1 circuit is simulated at some stage in the branching program. In the case of a QNC^1 circuit, the state of a wire is given by a qubit, which is, in general, entangled with the qubits carried by the other wires in the circuit. Therefore the state of this wire cannot be expressed in isolation, and there appears to be no alternative to creating that entangled state as part of any simulation. Thus the entire approach breaks down. One way to carry out such a construction, might be to apply a superposition of operations at each step: this extends the state space of the quantum computer and effectively provides many more clean qubits, making the model meaningless. Moreover all proposed implementations of quantum computation involve a classical, time-varying sequence of operations, applied to a quantum register. Since the control is classical, in any oblivious simulation the entangled quantum state of the simulated circuit must be encoded within the quantum register.

4 Limits on Computability

4.1 Requirements for one-clean-qubit simulation

We now turn to the main question of this paper: given a general m-qubit quantum circuit $\mathcal{C}$ that starts in the ideal $|00\ldots0\rangle$ starting state, can we simulate it by an n-qubit circuit $\mathcal{C}'$ in the one-clean-qubit model? We seek an *oblivious* simulation that works for any circuit $\mathcal{C}$. In an oblivious simulation, we must encode the 2^m basis states of $\mathcal{C}$ as distinguishable states of the n qubit register of $\mathcal{C}'$. Recall that the general state of $\mathcal{C}'$ is a density matrix corresponding to a subspace of dimension 2^{n-1}. There are two requirements for this encoding.

1. **Distinguishability.** The states must be distinguishable in the following sense: since we can prepare several copies of any state by repeating the simulation, we only require that there be a sequence of measurements on $O(\text{poly}(n))$ many copies of the state, that (with high probability) uniquely identify the state. Indeed, it is possible to do this with $n = m$, as follows: take the subspaces spanned by the basis vectors in the sets $A_b = \{x \in \{0,1\}^n : x \cdot b = 0 \mod 2\}$ for $b \in \{0,1\}^n$.

Let ρ_b be the mixed state over A_b. Measuring it gives a random vector x such that $x \cdot b = 0$. Repeating this n times gives n such vectors x_1, x_2, ..., x_n. With high probability, there is only one b such that $x_i \cdot b = 0$ for all i and it can be found by solving the system of linear equations $x_i \cdot b = 0$ with standard methods. (This is similar to what happens in Simon's quantum algorithm [19].)

2. **Permutability.** The quantum circuit $\mathcal{C}$ might carry out any unitary operation on its quantum state, and in particular an arbitrary permutation on its classical states. Therefore, we need to be able to carry out any permutation of states ρ_b by a unitary transformation. Unfortunately, this is not possible for the distinguishable set of states ρ_b described above.

   It is not hard to demonstrate an efficient encoding that satisfies this permutability condition, without distinguishability: take the subspaces spanned by the basis vectors in the sets $A_b = \{x = (x_1 \dots x_n) \in \{0,1\}^n : x_1 = 0 \text{ or } (x_2 \dots x_n) = b\}$ for $b \in \{0,1\}^{n-1}$ and let ρ_b be a mixed state over A_b.

   Those subspaces A_b (and the states ρ_b) can be permuted by unitary transformations that permute the vectors $|1x_2 \dots x_n\rangle$. However, they are not distinguishable. (Any measurement gives distributions that differ only by an exponentially small amount.)

We show that it is not possible to construct an efficient encoding that satisfies both conditions simultaneously. We do this first for classical encodings, then for quantum.

4.2 Classical impossibility result

To explain the challenge in the argument, let us first describe a simpler, classical analogue of our problem. We have a classical circuit taking inputs in $\{0,1\}^n$. This circuit is composed of reversible gates and executes a permutation of $\{0,1\}^n$. In its starting state, one bit is 0 and the rest are random. Thus, we have a uniform distribution over a set of size 2^{n-1} (all inputs with the first bit 0). Executing a reversible gate permutes $\{0,1\}^n$, mapping this uniform distribution to the uniform distribution over another set of size 2^{n-1}. The question is: how many sets $A_i \subset \{0,1\}^n$ of size 2^{n-1} can we have so that they are both distinguishable and permutable?

Theorem 2 *Let A_1, ..., A_M be permutable subsets of $\{0,1\}^n$ of size 2^{n-1}. If, for some i, j, $\frac{|A_i \cap A_j|}{|A_i|} \leq 1 - \epsilon$, then $M \leq \frac{2}{\epsilon} n$.*

Thus, if we require that the sets are distinguishable with probability $1/n^c$, we can only have $M = 2n^{c+1}$ sets. This means that the number of bits m that we can simulate is at most $\log M = (c+1)\log n + 1 = O(\log n)$.

Since the classical case is included only for guidance, we defer its proof to appendix 9.

4.3 Main theorem: quantum impossibility result

Let $M = 2^m$ be the total number of basis states of the ideal quantum computer being simulated. To simulate the computation, we encode each basis state by an arbitrary subspace X of dimension 2^{n-1} within the Hilbert space $\mathbb{C}^{2^n}$ of the computer. (The arbitrary subspaces replace the subsets A_i — equivalently axis-parallel subspaces — of the classical case.) The circuit of course can perform not just permutations of the basis, but general unitary operations. In sharp contrast with the classical case, two subspaces of half the dimensionality of the space typically will not intersect. Nevertheless, the large dimension of the subspaces imposes strict constraints on an operator which must permute them; the difficulty is in formulating the incompatibility of these requirements when the number of subspaces is large and the subspaces are required to be very distinct.

If the computer has k clean qubits, then the encodings are subspaces X of dimension 2^{n-k}. Let $\mathcal{X}$ be the set of these M subspaces $\{X\}$. We show that no oblivious simulation is possible unless n is exponentially larger than m or the number of clean qubits available is at least a constant fraction of m:

Theorem 3 *If $\mathcal{X}$ satisfies the following two constraints:*

1. *For every permutation π on M subspaces, there is a unitary transformation f_π such that $f_\pi(X) = \pi(X)$ for all $X \in \mathcal{X}$;*
2. *$\dim(X \cap Y)/\dim(X) < 1 - \frac{1}{\mathrm{poly}(m)}$ for any $X, Y \in \mathcal{X}$, $X \neq Y$;*

then $m \leq (2k + \log n)(1 + o(1))$.

The first constraint is permutability, and the second follows from distinguishability, as shown in:

Lemma 1 *Let X and Y be subspaces of dimension 2^{n-k} and let ρ_X and ρ_Y be the corresponding density matrices. If $\dim X \cap Y/\dim X = 1 - \epsilon$, then applying any measurement to ρ_X and ρ_Y results in output distributions which differ in variation distance by at most 2ϵ.*

Proof: Denote $Z = X \cap Y$. Let X' and Y' be the orthogonal complements of Z in X and Y; note $\epsilon = (\dim X')/(\dim X) = (\dim Y')/(\dim Y)$. Also,

$$\rho_X = (1-\epsilon)\rho_Z + \epsilon\rho_{X'}, \qquad \rho_Y = (1-\epsilon)\rho_Z + \epsilon\rho_{Y'}.$$

For any measurement on a density matrix $(1-\epsilon)\rho_1 + \epsilon\rho_2$, the output distribution is the convex combination (with coefficients $1-\epsilon, \epsilon$) of the output distributions for density matrices ρ_1 and ρ_2. The output distributions for ρ_X and ρ_Y are therefore within variation distance ϵ of that for ρ_Z. □

5 Proof outline

Recall that for every permutation π on the M subspaces in $\mathcal{X}$, there is a unitary transformation f_π such that $f_\pi(X) = \pi(X)$ for all $X \in \mathcal{X}$.

I. Suppose that the operators f_π form a group representation of S_M, i.e., $f_{\pi\sigma} = f_\pi f_\sigma$ for all π and $\sigma \in S_M$. Since 2^n is not much larger than M, this is a low dimensional representation of S_M. Assume further that the representation is irreducible. This implies a constraint on the Young diagram of this representation, namely that the first row or column must be very long (almost M). Fixing any $X \in \mathcal{X}$, we must be able to permute all other subspaces in $\mathcal{X}$ arbitrarily; this restricts our representation of S_M to a direct sum of irreps of S_{M-1}, each obtained by deleting an inside corner of the original Young diagram. Due to the shape constraint, one of these irreps has an overwhelming share of the total dimension; all of the $X \in \mathcal{X}$ must intersect in this subspace, contradicting distinguishability. This forms case 1 of the proof.

In case 2 we handle the situation in which the operators f_π still form a group representation of S_M, but this representation is reducible.

II. Actually the situation is more complicated.

Since $f_\pi f_\sigma(X) = f_\pi(\sigma(X)) = \pi\sigma(X)$, the transformation $f_\pi f_\sigma$ permutes the subspaces in the same way as $f_{\pi\sigma}$. However, this does not mean that $f_\pi f_\sigma = f_{\pi\sigma}$. In other words, the operators f_π may not form a representation.

This can arise in several ways. Since every X is a subspace, $-f_\pi f_\sigma(X)$ is the same as $f_\pi f_\sigma(X)$. So, we might have $f_\pi f_\sigma = -f_{\pi\sigma}$ for some π and σ. In general, we can also have more complicated transformations g such that $g(X) = X$ for all $X \in \mathcal{X}$. Then, we might have $f_\pi f_\sigma = g f_{\pi\sigma}$ for some π and σ.

Case 3 of the proof handles the situation when $f_\pi f_\sigma \neq f_{\pi\sigma}$ for some π and $\sigma \in S_M$. Let G be the group of transformations that map every subspace $X \in \hat{\mathcal{X}}$ to itself. Then, $f_\pi f_\sigma f_{\pi\sigma}^{-1}$ is an element of G for any $\pi, \sigma \in S_M$. We

would like to modify f so that this element becomes the identity for all π and $\sigma \in S_M$. Then, $f_\pi f_\sigma = f_{\pi\sigma}$, i.e., f_π would form a representation of S_M and we would be able to analyse this representation similarly to the previous section.

If for each π we select any element of G, call it g_π, and form the transformation $f'_\pi = g_\pi f_\pi$, then these transformations still implement the same permutation π of $\hat{\mathcal{X}}$. What we will do is show how to select the g_π so that on each irrep of G, the f' compose as a representation, up to a phase factor: $f'_{\pi\sigma} = c_{\pi,\sigma,i} f'_\pi f'_\sigma$ for some unit $c_{\pi,\sigma,i} \in \mathbb{C}$.

The next step is eliminating the phase factors $c_{\pi,\sigma,i}$. This is done by lifting these transformations to a space of at most quadratically larger size, in which the phases cancel and the transformations compose as a representation.

Case 3 occupies the bulk of the technical work.

6 Proof of the main theorem: case 1

In this section we consider the case that f_π form a representation ($f_\pi f_\sigma = f_{\pi\sigma}$ for all π and σ) and this representation is irreducible.

Lemma 2 *Let V be an irreducible representation of S_M of dimension $N \leq 2^n$ and X be a subspace fixed by a subgroup isomorphic to S_{M-1}. Then, the dimension of X is either at least $(1 - \frac{2n}{M})N$ or at most $\frac{2n}{M}N$.*

Proof: Let λ be the Young diagram of V. We show that λ has either a large first row or a large first column.

Lemma 3 *Either the first row or the first column of λ is of length at least $M - n$.*

Proof: This follows from a theorem by Rasala [16]:

Theorem 4 *[16, pp.151-152]*

1. *Let $A \leq M/2$ and ρ be an irreducible representation of S_M such that the first row of the Young diagram of ρ is of length exactly $M - A$. Then,*

$$\dim \rho \geq \varphi_A(M)$$

where $\varphi_A(M) = \frac{M-2A+1}{M-A+1}\binom{M}{A}$ is the dimension of the irreducible representation corresponding to the partition $(M - A, A)$.

2. *If ρ is an irreducible representation with both the first row and the first column of length at most $M/2$, then*

$$\dim \rho \geq \varphi_{\lfloor M/2 \rfloor}(M).$$

This theorem means that any representation with the first row of the Young diagram having length at most $M - A$, $A \leq M/2$ has dimension at least

$$\min_{B: A \leq B \leq M/2} \varphi_B(M).$$

A simple calculation shows that this expression is minimized by $B = A$ if $A \leq M/2 - c\sqrt{M}$ for some constant c and $B = \lfloor M/2 \rfloor$ if $A > M/2 - c\sqrt{M}$.[1]

To deduce our lemma, assume that the Young diagram of an irreducible representation of S_M has both first row and column of length at most $M - A$. We show that $N \geq 2^A$. Consider two cases:

Case 1: $A \geq M/2 - c\sqrt{M}$.

Notice that $A \leq M - \sqrt{M}$ because otherwise the Young diagram would fit into a square with a side less than $\sqrt{M}$ and area less than M. Theorem 4 implies that

$$\dim \rho \geq \varphi_{\lfloor M/2 \rfloor}(M) = \frac{1}{M} \binom{M}{\lfloor M/2 \rfloor} = \Omega\left(\frac{2^M}{M\sqrt{M}}\right) = 2^{M - \frac{3 \log M}{2} - O(1)} > 2^A.$$

Case 2: $A \leq M/2 - c\sqrt{M}$. Then,

$$\dim \rho \geq \varphi_A(M) = \frac{M - 2A + 1}{M - A + 1} \binom{M}{A} \geq \frac{1}{M} \left(\frac{M}{A}\right)^A$$

$$= \frac{1}{M} \left(\frac{M}{2A}\right)^A 2^A \geq 2^A.$$

□

Another lower bound on the longest row or column of a low dimension representation (for a different range of parameters) was given by Mischenko [11].

As described in section 2.4, V decomposes into irreducibles of S_{M-1} in the following way:

$$V = \bigoplus_{\lambda^-} \rho_{\lambda^-}$$

[1] In the second case, the lowest-dimensional representation actually has first row shorter than $M - A$. This is quite surprising because, in most cases, removing a square from the first row of a Young diagram and adding a square somewhere else increases the dimension.

where λ^- ranges over all shapes of size $M-1$ that can be obtained by deleting an "inside corner" from λ. If λ has a long first column or row, one of the λ^- has dimension that is almost the same as the dimension of λ.

Lemma 4 *Suppose the shape λ has a first row (column) of length $|\lambda| - \ell$ for $\ell < |\lambda|/2$. Let λ_1 denote the "λ^-" obtained by deleting the last element of the first row (column). Then* $\dim(\rho_{\lambda_1}) \geq \frac{|\lambda|-2\ell}{|\lambda|} \dim(\rho_\lambda)$.

Proof: Consider the ratio $\frac{\dim \rho_{\lambda_1}}{\dim \rho_\lambda} = \frac{1}{|\lambda|} \frac{\prod_{x\in\lambda} l(h_x)}{\prod_{x\in\lambda_1} l(h_x)}$. In the last ratio, points x outside of the first row (column) appear identically in the numerator and denominator. Moreover for each x in the first row (column) in the numerator other than the very last point (which contributes a factor of 1 in the numerator and is absent in the denominator), the ratio between its contributions in the numerator and denominator is $\frac{l(h_x)}{l(h_x)-1}$ where $l(h_x)$ is the length of its hook in λ.

Since there are only l squares below the first row, at most l squares in the first row have another square below them. Thus, none of $(|\lambda|-l)-l = |\lambda|-2l$ rightmost squares in the first row of λ has another square below it. The rightmost of them is absent in λ_1, the others have hook-lengths equal to 2, 3, ..., $|\lambda| - 2l$ in λ and 1, 2, ..., $|\lambda| - 2l - 1$ in λ_1. This lower-bounds the ratio by $\prod_{i=1}^{|\lambda|-2\ell-1} \frac{i+1}{i} = |\lambda| - 2\ell$. Therefore, $\frac{\dim \rho_{\lambda_1}}{\dim \rho_\lambda} \geq \frac{|\lambda|-2\ell}{|\lambda|}$. □

Let $W = \rho_{\lambda_1}$. Since X and W are invariant under S_{M-1}, $W \cap X$ must be invariant as well. Since W is irreducible, $W \cap X$ must be either 0 or W. If $W \cap X = W$, then $W \subseteq X$, $\dim X \geq \dim W \geq \dim W \geq (1 - \frac{2n}{M})N$ by Lemma 4. If $W \cap X = 0$, then $\dim X + \dim W \leq N$ and Lemma 4 implies $\dim X \leq N - \dim W \leq \frac{2n}{M} N$. □

7 Proof of the main theorem: case 2

In this section we consider the case that f_π form a representation ($f_\pi f_\sigma = f_{\pi\sigma}$ for all π and σ) but this representation is not irreducible.

In case 1, it was not possible to have a subspace X of dimension $(1/2)\dim V$ that is fixed by S_{M-1}. If we do not require the representation to be irreducible, there are many ways to construct such a subspace X (and a full set $\mathcal{X}$ of M permutable subspaces). However, in any such construction, any two of the subspaces must have a large-dimensional intersection.

Lemma 5 *Let $N < 2^n$. Let f_π be an N-dimensional representation of S_M that acts as a permutation representation on a collection of M subspaces $\hat{\mathcal{X}}$.*

Then, for any $X, Y \in \hat{\mathcal{X}}$,

$$\dim X - \dim X \cap Y \leq \frac{4n}{M} N.$$

Proof:

We decompose V into irreducible representations of S_M:

$$V = \oplus_{i=1}^{p} W_i \otimes E_i$$

where the W_i are nonisomorphic, and each E_i is a direct sum of several copies of the trivial representation $\mathbf{1}$.

We also decompose X into irreducibles of S_{M-1}^X (the subgroup of S_M that fixes X) in a similar way: $X = \oplus_{i=1}^{r} U_i \otimes F_i$ where U_i are distinct irreducibles of S_{M-1}^X and each F_i is a direct sum of several copies of the trivial representation $\mathbf{1}$.

Lemma 6 *For each* U_i*, there is at most one* j *such that* U_i *is the largest irreducible of* S_{M-1} *contained in* W_j*.*

Proof: By Lemma 4, the largest irreducible of S_{M-1} contained in W_j is the one with the Young diagram obtained by deleting the last square from the Young diagram of W_j. If W_j and $W_{j'}$ are different irreducibles of S_M, they must have different Young diagrams λ_j and $\lambda_{j'}$. Then, the diagrams obtained by deleting the last square of the first row from λ_j and $\lambda_{j'}$ are different as well. □

We reorder U_i so that, for each $i \in \{1, \ldots, p\}$, U_i is the largest irreducible of W_i. If there are S_{M-1}^X-irreducibles U_i that are not equal to the largest irreducible of any W_i ($i \in \{1, \ldots, p\}$), we denote them U_{p+1}, U_{p+2}, ...

Consider the intersection of $U_i \otimes F_i$ and $W_i \otimes E_i$. This intersection must be of the form $U_i \otimes K_i$ for some K_i. Let L_i be the complement of K_i in F_i. (If $i > p$, then $K_i = \{\overrightarrow{0}\}$ and $L_i = F_i$.) Define

$$X' = \sum_{i=1}^{p} U_i \otimes K_i, \qquad\qquad X'' = \sum_{i=1}^{r} U_i \otimes L_i.$$

Then $X = X' \oplus X''$. Let $\pi \in S_M$ be such that $\pi(X) = Y$. Define $Y' = f_\pi(X')$, $Y'' = f_\pi(X'')$. Then, $Y = Y' \oplus Y''$. The lemma 5 follows from two following lemmas:

Lemma 7

$$\dim(X') - \dim(X' \cap Y') \leq \frac{2n}{M} N.$$

Proof: We have $Y' = \oplus_{i=1}^{p} f_\pi(U_i) \otimes K_i$. By Lemma 4, the dimension of U_i (and $f_\pi(U_i)$) is at least $\frac{M-2n}{M} \dim(W_i)$. Therefore, $\dim(U_i) - \dim(U_i \cap f_\pi(U_i)) \leq \frac{2n}{M} \dim(W_i)$. The lemma follows by summation over all i. □

Lemma 8

$$\dim(X'') \leq \frac{2n}{M} N.$$

Proof: First, notice that $U_i \otimes L_i$ is orthogonal to $\sum_{j=1}^{p} U_j \otimes E_j$. (It is orthogonal to $U_i \otimes E_i$ because $(U_i \otimes E_i) \cap (U_i \otimes F_i) = U_i \otimes K_i$ and L_i is the complement of K_i in F_i. It is orthogonal to $U_j \otimes E_j$ for $j \neq i$ because U_i and U_j are different irreducibles of S_{M-1} (Lemma 6).

Therefore, $U_i \otimes L_i$ is contained in $\sum_{j=1}^{p} (W_j - U_j) \otimes E_j$ and X'' (which is the direct sum of $U_i \otimes L_i$ over all i) is contained in $\sum_{j=1}^{p} (W_j - U_j) \otimes E_j$ as well. We have

$$\dim X'' \leq \sum_{j=1}^{p} (\dim W_j - \dim U_j) \dim E_j \leq \frac{2n}{M} N,$$

with the last inequality following from Lemma 4. □

If f_π form a representation, Lemma 5 almost immediately implies Theorem 3. Namely, we have

$$\frac{\dim X \cap Y}{\dim X} = \frac{\dim X - (\dim X - \dim X \cap Y)}{\dim X}$$

$$\geq \frac{2^{n-k} - \frac{4n}{M} 2^n}{2^{n-k}} = 1 - \frac{2^{k+2} n}{M}.$$

If this is at most $1 - \frac{1}{\mathrm{poly}(m)}$, then $m \leq (k + \log n)(1 + o(1))$.

8 Proof of the main theorem: case 3

In this section we consider the case that f_π do not form a representation ($f_\pi f_\sigma \neq f_{\pi\sigma}$ for some π and σ).

8.1 Proof outline

Let G be the group of transformations that map every subspace $X \in \hat{\mathcal{X}}$ to itself. Then, $f_\pi f_\sigma f_{\pi\sigma}^{-1}$ is an element of G for any $\pi, \sigma \in S_M$. We would like to modify f so that this element becomes the identity for all π and $\sigma \in S_M$. Then, $f_\pi f_\sigma = f_{\pi\sigma}$, i.e., f_π would form a representation of S_M and we would be able to analyse this representation similarly to the previous section.

To achieve this, we look at $\mathbb{C}^{2^n}$ as a representation of G and express $\mathbb{C}^{2^n}$ as $V_1 \oplus V_2 \ldots \oplus V_k$, with V_i corresponding to different types of irreducible representations of G.

Then, we compose each f_π with an appropriate $g_\pi \in G$. The resulting transformation $f'_\pi = g_\pi f_\pi$ still implements the same permutation π of $\hat{\mathcal{X}}$ because g_π maps every $X \in \hat{\mathcal{X}}$ to itself. We can choose the transformations g_π so that, on every V_i, $f'_\pi f'_\sigma$ is the same as $f'_{\pi\sigma}$ up to a phase ($f'_{\pi\sigma} = c_{\pi,\sigma,i} f'_\pi f'_\sigma$ for some unit $c_{\pi,\sigma,i} \in \mathbb{C}$).

The next step is eliminating the phase factors $c_{\pi,\sigma,i}$. This is done by considering a larger space $V_1 \otimes V_1^* + \ldots + V_k \otimes V_k^*$ and transformations $f''_\pi = f'_\pi \otimes (f'_\pi)^*$ on this larger space. Then, the phase factors $c_{\pi,\sigma,i}$ (from f') and $c^*_{\pi,\sigma,i}$ (from $(f')^*$) cancel out and we get $f''_{\pi\sigma} = c_{\pi,\sigma,i} c^*_{\pi,\sigma,i} f''_\pi f''_\sigma = f''_\pi f''_\sigma$. Thus, f''_π form a representation of S_M on the linear space $V_1 \otimes V_1^* + \ldots + V_k \otimes V_k^*$. This representation can be then analysed similarly to case 2.

8.2 Representation up to phases $c_{\pi,\sigma,i}$

Let G be the group of unitary transformations that fix every one of the subspaces $X \in \hat{\mathcal{X}}$.

Then, $\mathbb{C}^{2^n}$ is a representation of G and all $X \in \hat{\mathcal{X}}$ are invariant subspaces. (They are fixed by every element of G according to the definition of G.) These invariant subspaces decompose into irreducible invariant subspaces.

Decompose $\mathbb{C}^{2^n}$ into irreducible invariant subspaces. Split those irreducible invariant subspaces into equivalence classes consisting of isomorphic irreducible subspaces. Let $E_1, \ldots, E_k$ be these equivalence classes. Let V_1 be the subspace of $\mathbb{C}^{2^n}$ spanned by all the irreducible subspaces in E_1 (i.e., the subspace spanned by all the vectors belonging to at least one subspace in E_1). Let V_2, ..., V_k be defined similarly.

Claim 1 *If $i, j \in \{1, \ldots, k\}$ and $i \neq j$, then $V_i \perp V_j$.*

Therefore, $\mathbb{C}^{2^n} = V_1 \oplus V_2 \oplus \ldots \oplus V_k$. Next, we show that transformations f_π map each V_i to some (possibly different) $V_{i'}$.

Claim 2 *Let V be an invariant subspace. Then, $f_\pi(V)$ is invariant as well. If V is irreducible, $f_\pi(V)$ is irreducible. Moreover, if V and V' are two isomorphic irreducible subspaces, $f_\pi(V)$ and $f_\pi(V')$ are isomorphic as well.*

Proof: The map $g \to f_\pi g f_\pi^{-1}$ is an automorphism of G. If V is invariant under the action of g, $f_\pi(V)$ is invariant under the action of $f_\pi g f_\pi^{-1}$. Therefore, if V is invariant under G, so is $f_\pi(V)$.

If $f_\pi(V)$ is not irreducible, it decomposes into two or more invariant subspaces: $f_\pi(V) = W_1 \oplus W_2$. Then, $f_\pi^{-1}(W_1)$ is invariant as well, implying that V is not irreducible.

Finally, let $h : V \to V'$ be a G-isomorphism of V and V' (an isomorphism that commutes with the action of G). Let $h' : f_\pi(V) \to f_\pi(V')$ be defined by $h' = f_\pi h f_\pi^{-1}$. Then, for any $g = f_\pi g' f_\pi^{-1}$, we have

$$h'g = (f_\pi h f_\pi^{-1})(f_\pi g' f_\pi^{-1}) = f_\pi h g' f_\pi^{-1} = f_\pi g' h f_\pi^{-1} = gh'$$

and every $g \in G$ can be expressed in the form $f_\pi g' f_\pi^{-1}$. Therefore, h' is a G-isomorphism of $f_\pi(V)$ and $f_\pi(V')$. □

Remark. V does not have to be isomorphic to $f_\pi(V)$ as a representation of G. f_π establishes the isomorphism of g on V with $f_\pi g f_\pi^{-1}$ on $f_\pi(V)$, but $f_\pi g f_\pi^{-1}$ does not have to equal g on $f_\pi(V)$.

Claim 3 *For every $i \in \{1, \ldots, k\}$ there is an i' such that $f_\pi(V_i) = V_{i'}$.*

Proof: By Claim 2, every two isomorphic irreducible subspaces get mapped to isomorphic irreducible subspaces. Therefore, all subspaces in E_i get mapped to subspaces in the same $E_{i'}$ and $f_\pi(V_i) \subseteq V_{i'}$. Similar reasoning applied to f_π^{-1} implies $f_\pi^{-1}(V_{i'}) \subseteq V_i$. □

For each $i \in \{1, \ldots, k\}$, V_i is the direct sum of some number of isomorphic irreducible subspaces: $V_i = V_{i1} \oplus V_{i2} \oplus \ldots V_{ij_i}$. We fix G-isomorphisms $h_{ijj'}$ between V_{ij} and $V_{ij'}$ so that $h_{ij'j''} h_{ijj'} = h_{ijj''}$.

Lemma 9 *[5, Schur's lemma] Let V and V' be an isomorphic irreducible representations of G. Then, for any two G-isomorphisms $h : V \to V'$ and $h' : V \to V'$, there is a constant c such that $h = ch'$.*

Therefore, each $h_{ijj'}$ is unique up to a multiplicative constant. The isomorphisms can be made to compose properly by adjusting these constants. Note that h_{ijj} is the identity.

Claim 4 *$W \subseteq V_i$ is an irreducible invariant subspace if and only if*

$$W = \{a_j x + a_{j+1} h_{ij(j+1)}(x) + \ldots + a_{j_i} h_{ijj_i}(x) | x \in V_{ij}\}$$

for some $j \in \{1, \ldots, j_i\}$ and $a_j, \ldots, a_{j_i} \in \mathbb{C}$.

Proof: "If" part:

Invariance:

$$g(\sum_{\ell=j}^{j_i} a_\ell h_{ij\ell}(x)) = \sum_{\ell=j}^{j_i} a_\ell g(h_{ij\ell}(x)) = \sum_{\ell=j}^{j_i} a_\ell h_{ij\ell}(g(x))$$

because each $h_{ij\ell}$ is a G-isomorphism.

W is irreducible because, if $W_1 \subset W$ and W_1 is invariant, then

$$\{x | a_j x + a_{j+1} h_{ij(j+1)}(x) + \ldots + a_{j_i} h_{ijj_i}(x) \in W_1\}$$

is an invariant subspace of V_{ij}; but V_{ij} is irreducible and $\dim(W) = \dim(V_{ij})$.

"Only if" part:

Let W be an irreducible invariant subspace of V_i. Let $x' \in W$. Then, we can write x' as $x'_1 + \ldots + x'_{j_i}$, $x'_1 \in V_{i1}$, ..., $x'_{j_i} \in V_{ij_i}$. If $x' \neq x'' \in W$, then for any index j, $x'_j \neq x''_j$ or $x'_j = x''_j = 0$. (Otherwise, $W \cap \bigoplus_{\ell \neq j} V_{i\ell}$ is a nontrivial subspace of W. It is invariant because V_i and V_{ij} are invariant. Contradiction with the irreducibility of W.)

Let j be the smallest index for which there is an $x' \in W$ with $x'_j \neq 0$. Then, for every $x \in V_{ij}$, there is an $x' \in W$ with $x'_j = x$. (For, if A and B are invariant subspaces of a unitary representation, the projection of A onto B is invariant. Apply this with $A = W$ and $B = V_{ij}$, then use the irreducibility of V_{ij}.)

The above considerations allow us to define the mapping $h_{jj'} : V_{ij} \to V_{ij'}$ by $h_{jj'}(x'_j) = x'_{j'}$. By the definition, $h_{jj'}(g(x'_j)) = h_{jj'}((g(x'))_j) = (g(x'))_{j'} = g(x'_{j'}) = g(h_{jj'}(x'_j))$, so $h_{jj'}$ is a G-isomorphism. By Schur's lemma, this implies that $h_{jj'} = a_{j'} h_{ijj'}$ for some $a_{j'} \in \mathbb{C}$. $\square$

In general, if W is any irreducible invariant subspace of $\mathbb{C}^{2^n}$, then W must be in the form described by claim 4 for some i. (W belongs to some equivalence class E_i and therefore is contained in the corresponding V_i.)

For each V_{i1} ($i \in \{1, \ldots, k\}$), we fix an orthonormal basis $v_{i1}, \ldots, v_{it}$. This also fixes a related basis $h_{i1j}(v_{i1})$, ..., $h_{i1j}(v_{it})$ for each V_{ij}. Moreover, we also get a similar basis

$$\sum_{\ell=j}^{j_i} a_\ell h_{i1\ell}(v_{i1}), \ldots, \sum_{\ell=j}^{j_i} a_\ell h_{i1\ell}(v_{it})$$

for every invariant irreducible $W \subseteq V_i$ because any such W can be written in the form given by claim 4. We call these bases *designated*.

This designated basis is exactly the basis for W that can be obtained by applying the isomorphism between V_{i1} and W to the basis for V_{i1}. Moreover,

if W, W' are two isomorphic irreducible subspaces, the designated basis for W is mapped to the designated basis for W' by the isomorphism between W and W'.

We are going to impose the following condition on f'_π:

Condition. Let W be an irreducible representation of G and $w_1, \ldots, w_l$ be the designated basis of W. Let $w'_1, \ldots, w'_l$ be the designated basis of $f'_\pi(W)$. Then, there exists $c \in \mathbb{C}$, $|c| = 1$ such that $f'_\pi(w_1) = cw'_1$, ..., $f'_\pi(w_l) = cw'_l$.

Next, we show that this condition suffices to guarantee $f'_\pi f'_\sigma = c_{\pi,\sigma,i} f'_{\pi\sigma}$ on every V_i and that any f_π that permutes $X \in \hat{\mathcal{X}}$ without satisfying this condition can be transformed into f'_π that satisfies the condition and still permutes the subspaces in the same way.

First, we show that it is enough to ensure that the designated basis of V_{i1} is mapped correctly for every $i \in \{1, \ldots, k\}$.

Claim 5 *Assume that the condition is true for $W = V_{i1}$. Then, it is also true for any irreducible $W \subseteq V_i$.*

Proof: Let h be the isomorphism between V_{i1}, W. Note that h maps the designated basis of V_{i1} to the designated basis of W.

Then (by claim 2) $f'_\pi h (f'_\pi)^{-1}$ is an isomorphism between $f'_\pi(V_{i1})$ and $f'_\pi(W)$. We know that there is an isomorphism between these two irreducibles that maps the designated basis of one of them to the designated basis of the other. By Schur's lemma, any two isomorphisms of irreducible subspaces can differ only by a multiplicative constant c. The unitarity of $f'_\pi h (f'_\pi)^{-1}$ implies that $|c| = 1$.

Therefore, $f'_\pi h (f'_\pi)^{-1}$ maps the designated basis of $f'_\pi(V_{i1})$ to c times the designated basis of $f'_\pi h (f'_\pi)^{-1}(f'_\pi(V_{i1})) = f'_\pi h(V_{i1}) = f'_\pi(W)$. We know that $(f'_\pi)^{-1}$ maps the designated basis of $f'_\pi(V_{i1})$ to the designated basis of V_{i1} and that h maps the designated basis of V_{i1} to the designated basis of W. This implies that f'_π maps the designated basis of W to c times the designated basis of $f'_\pi(W)$. □

Next, we show how to transform f_π into f'_π that performs the same permutation π of $\hat{\mathcal{X}}$ and maps the designated basis of every V_{i1} as required.

Let W_1, ..., W_k be $f_\pi(V_{11})$, ..., $f_\pi(V_{k1})$. Each of W_i lies within one of $V_1, \ldots, V_k$. Denote this subspace $V_{i'}$. Then, for $i \neq j$, $V_{i'} \neq V_{j'}$. For each $i \in \{1, \ldots, k\}$, we define a unitary transformation $g_{\pi,i}$ on V_i such that $g_{\pi,i'} f_\pi$ maps the designated basis of V_{i1} to the designated basis of $f_\pi(V_{i1})$.

By Claim 4, the irreducible subspace $W_i = f_\pi(V_{i1})$ is just

$$\{a_{i'j}x + a_{i'(j+1)}h_{i'j(j+1)}(x) + \ldots + a_{i'j_{i'}}h_{i'jj_{i'}}(x) | x \in V_{i'j}\}$$

for some j. Moreover, the mapping that maps each $v \in W_i$ to its $V_{i'j}$-component is an isomorphism of W_i and $V_{i'j}$ w.r.t. G (similarly to proof of Claim 4).

Let $v_1, \ldots, v_l$ be the designated basis of V_{i1}, $v'_1, \ldots, v'_l$ be $f_\pi(v_1), \ldots, f_\pi(v_l)$ and v''_1, ..., v''_l be the $V_{i'j}$ components of v'_1, ..., v'_l.

Let w_1, ..., w_l be the designated basis of $V_{i'j}$ and $g_{\pi,i'j}$ be the unitary transformation on $V_{i'j}$ that maps $v''_1, \ldots, v''_l$ to w_1, ..., w_l. We define a unitary transformation $g_{\pi,i'j'}$ (for every $j' \neq j$) on $V_{i'j'}$ to be $h_{i'jj'} g_{\pi,i'j} h^{-1}_{i'jj'}$. Finally, we take the transformation $g_{\pi,i'}$ of $V_{i'}$ that is equal to $g_{\pi,i'j}$ on each $V_{i'j}$. Then, $g_{\pi,i'}$ maps

$$v'_1 = a_{i'j} v''_1 + a_{i'(j+1)} h_{i'j(j+1)}(v''_1) + \ldots + a_{i'j_{i'}} h_{i'jj_{i'}}(v''_1)$$

to

$$\begin{aligned} a_{i'j} g_{\pi,i'j}(v''_1) + a_{i'(j+1)} g_{\pi,i'(j+1)} h_{i'j(j+1)}(v''_1) + \ldots = \\ a_{i'j} g_{\pi,i'j}(v''_1) + a_{i'(j+1)} h_{i'j(j+1)} g_{\pi,i'j}(v''_1) + \ldots \\ = a_{i'j} w_1 + a_{i'(j+1)} h_{i'j(j+1)}(w_1) + \ldots \end{aligned}$$

which is exactly the first vector of the designated basis for W_i. The same is true for v'_2, ..., v'_l, implying that $g_{\pi,i'} f_\pi$ maps the designated basis of V_{i1} to the designated basis of W_i.

Now, we take g_π that is equal to $g_{\pi,i}$ on each V_i and take $f'_\pi = g_\pi f_\pi$.

Claim 6 *g_π preserves all $X \in \hat{\mathcal{X}}$.*

Proof: By definition, the restriction $g_\pi|_{V_i}$ is equal to $g_{\pi,i}$, and $g_{\pi,i}$ clearly preserves $V_{i1}, \ldots, V_{ij_i}$. Moreover, $g_{\pi,i}$ (and, hence, g_π) preserves any irreducible subspace $W \subseteq V_i$ because any such subspace is in the form of claim 4.

Every $X \in \hat{\mathcal{X}}$ is invariant under G. Therefore, it decomposes into a direct sum of irreducible subspaces. Each of these subspaces is in one of the classes E_1, ..., E_k and, therefore, lies in one of V_1, ..., V_k. This means that it is preserved by g_π. Therefore, X which is a direct sum of such irreducible subspaces is preserved by g_π as well. □

Hence, $f'_\pi = g_\pi f_\pi$ realizes the same permutation π of $X \in \hat{\mathcal{X}}$ as f_π.

Claim 7 *On every V_i, $f'_\pi f'_\sigma = c_{\pi,\sigma,i} f'_{\pi\sigma}$ for some $c_{\pi,\sigma,i} \in \mathbb{C}$.*

Proof: This is equivalent to showing that $(f'_{\pi\sigma})^{-1} f'_\pi f'_\sigma$ is equal to $c_{\pi,\sigma,i}$ times the identity. To show that, notice that $(f'_{\pi\sigma})^{-1} f'_\pi f'_\sigma$ maps every subspace

$X \in \hat{\mathcal{X}}$ to itself because $(f'_{\pi\sigma})^{-1}$ performs the inverse of the permutation $\pi\sigma$ on $\hat{\mathcal{X}}$. Therefore, $(f'_{\pi\sigma})^{-1} f'_\pi f'_\sigma \in G$. This means that V_{ij} are all preserved by $(f'_{\pi\sigma})^{-1} f'_\pi f'_\sigma$.

Moreover, f'_σ, f'_π and $f^{-1}_{\pi\sigma}$ all map the designated bases to c-times designated bases (Claim 5). Therefore, $(f'_{\pi\sigma})^{-1} f'_\pi f'_\sigma$ maps the designated basis of V_{ij} to c times the designated basis of $(f'_{\pi\sigma})^{-1} f'_\pi f'_\sigma(V_{ij}) = V_{ij}$.

It remains to show that c is the same for all irreducible subspaces V_{ij} contained in V_i. Let c_j and $c_{j'}$ be the values of c for V_{ij} and $V_{ij'}$. Consider the subspace

$$W = \{x + h_{ijj'}(x) | x \in V_{ij}\}.$$

By Claim 4, this is an irreducible invariant subspace. Now, $(f'_{\pi\sigma})^{-1} f'_\pi f'_\sigma$ maps it to

$$W' = \{c_j x + c_{j'} h_{ijj'}(x) | x \in V_{ij}\} =$$

$$\{x + \frac{c_j}{c_{j'}} h_{ijj'}(x) | x \in V_{ij}\}.$$

The invariance of W means that $W' = W$ and $c_j = c_{j'}$.

Therefore, c_j are all equal. This means that $(f'_{\pi\sigma})^{-1}(x) = c_j x$ for all $x \in V_i$ because the designated bases of V_{ij} together form a basis for entire subspace V_i. $\square$

Unfortunately, arguments of this type (composing f_π with an appropriate transformation that fixes all U_i) cannot be used to eliminate phases $c_{\pi,\sigma,i}$.

The reason for this is that there exist so-called *projective representations*. A projective representation is a set of maps f_π such that $f_\pi f_\sigma = c_{\pi,\sigma} f_{\pi\sigma}$, $c_{\pi,\sigma} \in \mathbb{C}$. It is known that the symmetric group has projective representations which are not equivalent to any of the usual representations[9].

One possible solution would be to use the standard forms of projective representations which are quite well studied[9]. However, to be able to use them, we would need to show that the multiplicative constants $c_{\pi,\sigma,i}$ are the same for all V_i (or show that we can split all V_i in several groups so that $c_{\pi,\sigma,i}$ is the same within one group) and we do not know if this is possible.

Our solution is to replace f'_π by transformations f''_π on a larger space $V_1 \otimes V_1^* + \ldots V_k \otimes V_k^*$ so that $f''_\pi f''_\sigma = f''_{\pi\sigma}$. Then, f''_π form a representation in the usual sense and we can analyse them similarly to section 7.

8.3 Fixing the phase problem

We split $V_1, \ldots, V_k$ into equivalence classes $\mathcal{V}_1, \ldots \mathcal{V}_l$. V_i and V_j are in one class if there is a $\pi \in S_M$ such that $f_\pi(V_i) = V_j$. Let W_i be the union of all V_j that belong to $\mathcal{V}_i$. Then, $f_\pi(W_i) = W_i$ for any $\pi \in S_M$ (because f_π

maps every $V_j \in \mathcal{V}_i$ to some $V_{j'} \in \mathcal{V}_i$). Therefore, we can look at each W_i separately.

Each subspace X is invariant under any transformation in G (because we defined G as the group of transformations that fix all X). Therefore, X is a direct sum of irreducible representations of G. Each of those irreducibles is contained in some V_i. Therefore,

$$X = \oplus_{i=1}^{k} X \cap V_i.$$

If we group together all V_i that belong to the same equivalence class $\mathcal{V}_i$, we get

$$X = \oplus_{j=1}^{l} X \cap W_j, \qquad X \cap W_j = \oplus_{i \in \mathcal{V}_j} X \cap V_i.$$

The intersections $X \cap Y$ decompose similarly. So, we can show the following lemma and then obtain Theorem 3 by summation over W_j.

Lemma 10 *Let $X, Y \in \hat{\mathcal{X}}$. Then, for any j,*

$$\dim X \cap W_j - \dim X \cap Y \cap W_j \leq \sqrt{\frac{4n}{M}} \dim W_j.$$

Proof: To simplify the notation, assume that $W_j = V_1 \oplus V_2 \ldots \oplus V_l$.

Consider the linear space $W_j' = V_1 \otimes V_1^* \oplus \ldots V_l \otimes V_l^*$ and the linear transformations $f_\pi'' = f_\pi' \otimes (f_\pi')^*$. These linear transformations form a representation because

$$f_{\pi\sigma}' \otimes (f_{\pi\sigma}')^* = c_{\pi,\sigma,i} f_\pi' f_\sigma' \otimes c_{\pi,\sigma,i}^* (f_\pi')^* (f_\sigma')^* =$$

$$f_\pi' f_\sigma' \otimes (f_\pi')^* (f_\sigma')^* = f_\pi'' f_\sigma''$$

on every $V_i \otimes V_i^*$. The counterpart of X is

$$X' = \oplus_{i=1}^{l} (X \cap V_i) \otimes (X \cap V_i)^*.$$

We have $f_\pi''(X') = Y'$. (To see this, consider one of $(X \cap V_i) \otimes (X \cap V_i)^*$. Assume that f_π' maps V_i to $V_{i'}$. Then, $f_\pi'(X) = Y$ implies $f_\pi'(X \cap V_i) = Y \cap V_{i'}$ and

$$f_\pi''((X \cap V_i) \otimes (X \cap V_i)^*) = (Y \cap V_{i'}) \otimes (Y \cap V_{i'})^*.$$

Combining these equalities for all V_i gives $f_\pi''(X') = Y'$.)

In particular, $f_\pi''(X') = Y'$ means that X' is invariant under all $\pi \in S_M$ satisfying $\pi(X) = X$. Therefore, by Lemma 5,

$$\dim X' - \dim X' \cap Y' \leq \frac{4n}{M} \dim W_j'. \tag{3}$$

We use this inequality to derive a bound on $\dim X \cap W_j - \dim X \cap Y \cap W_j$. To do this, we relate the dimensions of $X \cap V_i$ and $X' \cap (V_i \otimes V_i^*)$. Remember that

$$X \cap W_j = \oplus_{i=1}^{l} (X \cap V_i), \tag{4}$$

$$X \cap Y \cap W_j = \oplus_{i=1}^{l} (X \cap Y \cap V_i) \tag{5}$$

Let d_i and d_i' be the dimensions of $X \cap V_i$ and $X \cap Y \cap V_i$. Then, (4) and (5) imply that $\dim X \cap W_j = \sum_{i=1}^{l} d_i$, $\dim X \cap Y \cap W_j = \sum_{i=1}^{l} d_i'$ and

$$\dim X \cap W_j - \dim X \cap Y \cap W_j = \sum_{i=1}^{l} (d_i - d_i').$$

If we look at $V_i \otimes V_i^*$, then

$$X' \cap (V_i \otimes V_i^*) = (X \cap V_i) \otimes (X \cap V_i)^*.$$

This implies $\dim X' \cap (V_i \otimes V_i^*) = d_i^2$ and $\dim X' = \sum_i d_i^2$. Similarly, $\dim X' \cap Y' = \sum_{i=1}^{l} {d_i'}^2$.

Let d be the dimension of V_1. Then, the dimensions of $V_2, \ldots, V_l$ are d as well because, for every $i \in \{2, \ldots, l\}$, there is a unitary f_π such that $f_\pi(V_1) = V_i$. Therefore, $\dim W_j = ld$. Also, $\dim W_j' = ld^2$ because $\dim V_i \otimes V_i^* = d^2$ for every $i \in \{1, \ldots, l\}$. Hence, we have

$$\frac{\dim X \cap W_j - \dim X \cap Y \cap W_j}{\dim W_j} = \frac{\sum_{i=1}^{l} (d_i - d_i)}{ld} =$$

$$\frac{1}{l} \sum_{i=1}^{l} \sqrt{\frac{(d_i - d_i')^2}{d^2}} \leq \frac{1}{l} \sum_{i=1}^{l} \sqrt{\frac{(d_i - d_i')(d_i + d_i')}{d^2}} =$$

$$\frac{1}{l} \sum_{i=1}^{l} \sqrt{\frac{(d_i^2 - {d_i'}^2)}{d^2}}.$$

The convexity of the square root implies that this is at most

$$\sqrt{\frac{\sum_{i=1}^{l} (d_i^2 - {d_i'}^2)}{ld^2}} = \sqrt{\frac{\dim X' - \dim X' \cap Y'}{\dim W_j'}}.$$

Equation (3) implies that this is at most $\sqrt{4n/M}$. This completes the proof of lemma. □

By summing over W_j's, we get

$$\dim X - \dim X \cap Y \leq \sqrt{\frac{4n}{M}} \sum_j \dim W_j = \sqrt{\frac{4n}{M}} 2^n$$

and the rest is similar to the end of section 7. This completes the proof of Theorem 3. □

References

[1] Ablyev F., Moore C., Pollett C., Quantum and stochastic branching programs of bounded width. *Proceedings of ICALP 2002*, 343-354.

[2] Barrington, D. A, "Bounded-width polynomial-size branching programs recognize exactly those languages in NC1", *J. Computer and System Sciences* 38(1):150-164, February 1989.

[3] Chuang, I. L., Vandersypen, L. M. K., Zhou, X., Leung D. W. and Lloyd, S., "Experimental realization of a quantum algorithm", *Nature* 393, 143-146, 1998.

[4] Cory, D. G., Fahmy, A. F. and Havel, T. F., "Ensemble Quantum Computing by Nuclear Magnetic Resonance Spectroscopy", *Proc. Natl. Acad. Sci.* 94:1634-1639, 1997.

[5] Fulton, W. and Harris, J., "Representation Theory. A First Course". Springer-Verlag, 1991.

[6] Gershenfeld, N. and Chuang, I., *Science* 275, pp. 350, 1997.

[7] M. Grigni, L. J. Schulman, M. Vazirani, U. Vazirani. Quantum mechanical algorithms for the hidden subgroup problem. *Proceedings of STOC'01*, pp. 68-74.

[8] S. Hallgren, A. Russell, A. Ta-Shma. Normal subgroup reconstruction and quantum computation using group representations. *Proceedings of STOC'00*, pp. 627-635.

[9] Hoffman, P. N. and Humphreys, J. F. "Projective Representations of the Symmetric Groups". Oxford University Press, 1992.

[10] Knill, E. and Laflamme, R., "On the power of one bit of quantum information", *Phys. Rev. Lett.* 81:5672, 1998.

[11] Mishchenko, S. P. "Lower bounds on the dimensions of irreducible representations of symmetric groups and of the exponents of the exponential of varieties of Lie algebras", *Matematicheskiĭ Sbornik*, 187:83-94, 1996.

[12] C. Moore, M. Nillson. Parallel quantum computation and quantum codes. *SIAM Journal on Computing*, to appear.

[13] Nielsen, M., "Probability distributions consistent with a mixed state", LANL archive http://xxx.lanl.gov/abs/quant-ph/9909020 .

[14] M. Nielsen, I. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 2000.

[15] S. Parker, M. Plenio. Efficient factorization with a single pure qubit and $\log N$ mixed qubits. *Physical Review Letters*, 85:3049-3052, 2000.

[16] Rasala, R., On the minimal degrees of characters of S_n. *Journal of Algebra*, 45:132-181, 1977.

[17] B. Sagan. *The Symmetric Group* (2nd edition). Springer, 2001.

[18] Schulman, L. J. and Vazirani, U. V., "Molecular Scale Heat Engines and Scalable Quantum Computation", *Proc. 31'st Ann. ACM Symp. Theory Comput.* 1999.

[19] D. Simon. On the power of quantum computation. *SIAM Journal on Computing*, 26:1474-1483, 1997.

[20] L. Vandersypen, M. Steffen, G. Breyta, C. Yannoni, R. Cleve, I. Chuang. Experimental Realization of an Order-Finding Algorithm with an NMR Quantum Computer. *Physical Review Letters*, 85:5452-5455, 2000.

[21] A. Verhulst, O. Liivak, M. Sherwood, H. Vieth, I. Chuang. Non-thermal nuclear magnetic resonance quantum computing using hyperpolarized Xenon. *Applied Physics Letters*, 79:2480, 2001. quant-ph/0105147.

9 Proof of the classical impossibility result

Proof of theorem 2:

Let $i \neq j$ and $k \neq l$. Then, there is a permutation π such that $\pi(i) = k$, $\pi(j) = l$. By permutability, this means that there is σ such that $\sigma(A_i) = A_k$

and $\sigma(A_j) = A_l$. Therefore, the intersections $A_i \cap A_j$ and $A_k \cap A_l$ must have the same number of elements. Thus, $|A_i \cap A_j| \leq (1-\epsilon)|A_i| = (1-\epsilon)2^{n-1}$ for every $i \neq j$. Equivalently, $|A_i - A_j| \geq \epsilon 2^{n-1}$.

Lemma 11 *Let $M > 2/\epsilon$. There exists $x \in \{0,1\}^n$ that belongs to k sets A_i for $\frac{\epsilon}{2}M \leq k \leq (1-\frac{\epsilon}{2})M$.*

Proof: We count the number of triples (i,j,x), $i,j \in \{1,\ldots,M\}$, $x \in \{0,1\}^n$ such that $i \neq j$, $x \in A_i$, $x \notin A_j$.

If $|A_i - A_j| \geq \epsilon 2^{n-1}$, there are at least $\epsilon 2^{n-1}$ x that belong to A_i but not to A_j. Since $A_i - A_j$ is of the same size for all i,j, $i \neq j$, this must be true for every pair i,j, $i \neq j$. Since there are $M(M-1)$ such pairs, we get at least $\epsilon M(M-1)2^{n-1}$ triples.

On the other hand, if x belongs to at least $(1-\frac{\epsilon}{2})M$ or at most $\frac{\epsilon}{2}M$ sets A_i, there are at most $\frac{\epsilon}{2}M(1-\frac{\epsilon}{2})M$ pairs (i,j) such that $x \in A_i$ but $x \notin A_j$. If this is true for every of 2^n possible $x \in \{0,1\}^n$, we get at most $\frac{\epsilon}{2}(1-\frac{\epsilon}{2})M^2 2^n$ triples. Then, we must have

$$\epsilon M(M-1)2^{n-1} \leq \frac{\epsilon}{2}(1-\frac{\epsilon}{2})M^2 2^n.$$

This only happens when $(M-1) \leq (1-\frac{\epsilon}{2})M$ which is equivalent to $\epsilon M \leq 2$. In this case, $M \leq \frac{2}{\epsilon}$ $\Box$

Let $x \in \{0,1\}^n$ belong to k sets A_i. W.l.o.g., assume that x belongs to A_1, A_2, ..., A_k. Let i_1, i_2, ..., i_k be k different numbers from $\{1,\ldots,M\}$. Take a permutation $\pi \in S_M$ that $\pi(1) = i_1$, ..., $\pi(k) = i_k$. By the permutability requirement, there is a $\sigma \in S_N$ such that $\sigma(A_i) = \sigma(A_{\pi(i)})$. This means that $\sigma(x)$ belongs to $A_{\pi(1)}$, ..., $A_{\pi(k)}$ and no other set. We have shown that for any i_1, ..., i_k, there is a $y \in \{0,1\}^n$ that belongs to A_{i_1}, A_{i_2}, ..., A_{i_k} and no other set. There are $\binom{M}{k}$ ways to choose i_1, i_2, ..., i_k. Therefore, there must be at least $\binom{M}{k}$ different $y \in \{0,1\}^n$. This means that

$$\binom{M}{k} \leq 2^n.$$

W.l.o.g., we can assume that $k \leq M/2$. (Otherwise, just replace k by $M-k$ and $\binom{M}{k}$ will stay the same.) We have $\binom{M}{k} \geq (\frac{M}{k})^k$. Since $k \leq M/2$, this is at least 2^k. Since $k \geq \frac{\epsilon}{2}M$, $2^k \geq 2^{\epsilon M/2}$. Thus, we have

$$2^{\epsilon M/2} \leq 2^n.$$

This implies $\epsilon M/2 \leq n$ and $M \leq \frac{2}{\epsilon}N$. $\Box$

The large size of the subsets means that we have far more constraints than we have degrees of freedom. The combinatorial argument shows that if the requirement of full permutability is imposed, and we have a super-polynomial (in n) number of subsets, then the difference of every two sets must be a vanishing fraction of the size of the sets. The two types of sets $\{A_b\}$ described above, however, separately achieve distinguishability and permutability.