Quantum states cannot be transmitted efficiently classically

Ashley Montanaro[*]

December 21, 2016

Abstract

We show that any classical communication protocol that can approximately simulate the result of applying an arbitrary measurement (held by one party) to a quantum state of n qubits (held by another) must transmit at least 2^n bits, up to constant factors. The argument is based on a lower bound on the classical communication complexity of a distributed variant of the Fourier sampling problem. We obtain two optimal quantum-classical separations as corollaries. First, a sampling problem which can be solved with one quantum query to the input, but which requires order-N classical queries for an input of size N. Second, a nonlocal task which can be solved using n Bell pairs, but for which any approximate classical solution must communicate 2^n bits, up to constant factors.

1 Introduction

How much classical information does it take to store or transmit a quantum state? In one sense, the answer is clear: a pure state of n qubits corresponds to a unit vector in $\mathbb{C}^{2^n}$, which requires $2^{n+1}-2$ real numbers to be specified exactly (2^n complex numbers, except that the absolute value of the last one is already determined by normalisation, and we can ignore an irrelevant overall phase). However, surprisingly, a number of known results suggest that the amount of classical information required to transmit a quantum state could actually be substantially less than this.

The most famous result of this nature is Holevo's Theorem [22], a corollary of which is that the classical information content of a quantum state of n qubits is bounded by n bits. A related result is a bound of Nayak [32] which implies that any quantum communication protocol which transmits n bits with success probability $\delta > 0$ must send at least $n - \log_2(1/\delta)$ qubits. Indeed, the number of qubits transmitted must be linear in n even if we seek only to retrieve one of the bits with high probability [4]. It was also shown by Aaronson [1] that to predict the outcomes of most measurements drawn from some probability distribution on an n-qubit state, one needs to make only $O(n)$ sample measurements drawn from the same distribution. This result motivated Aaronson to make the provocative statement that "while the effective dimension of an n-qubit Hilbert space *appears* to be exponential in n, in the sense that is relevant for approximate learning and prediction, this appearance is illusory".

One way to gain intuition for these results is to observe that having one copy of an n-qubit quantum state $|\psi\rangle$ does not allow the retrieval of up to around 2^n parameters precisely. For any measurement which we can perform on $|\psi\rangle$, if we instead applied it to $|\widetilde{\psi}\rangle \approx |\psi\rangle$, the distribution on measurement outcomes would be almost the same. Therefore, the most we can reasonably ask

[*] School of Mathematics, University of Bristol, UK; ashley.montanaro@bristol.ac.uk.

of a classical protocol designed to store or transmit $|\psi\rangle$ is to achieve what the quantum protocol itself does: allow approximate reproduction of any measurement which we could perform on $|\psi\rangle$.

We can express this task within the framework of a communication game. Imagine there are two parties (Alice and Bob), where Alice has a description of a pure quantum state $|\psi\rangle$ of n qubits, and Bob has a description of a quantum measurement (POVM) M. Let $p_M(|\psi\rangle)$ denote the probability distribution resulting from applying the measurement M to $|\psi\rangle$. Then we ask that the classical protocol allows Alice and Bob to sample from a distribution $\widetilde{p}_M(|\psi\rangle)$ such that $\|p_M(|\psi\rangle) - \widetilde{p}_M(|\psi\rangle)\|_1 \leq \epsilon$, for some $\epsilon > 0$, where $\|\cdot\|_1$ is the ℓ_1 distance $\|v\|_1 = \sum_i |v_i|$. Call this task *Quantum State Sampling*. Quantum State Sampling can clearly be solved with $\epsilon = 0$ by a quantum protocol communicating n qubits: Alice just sends $|\psi\rangle$ to Bob, who measures according to M.

Quantum State Sampling can also be solved with $O(2^n \log(1/\epsilon))$ bits of classical communication. It is sufficient for Alice to encode $|\psi\rangle$ using an ϵ-net with respect to the trace norm [21], i.e. a set of states $\{|\psi_i\rangle\}$ such that, for all $|\psi\rangle$, $\||\psi\rangle\langle\psi| - |\psi_i\rangle\langle\psi_i|\|_1 \leq \epsilon$ for some i. Then Alice sends the identity of the closest state in the ϵ-net to Bob, who samples from the distribution corresponding to applying M to that state. As there are ϵ-nets of size $(5/\epsilon)^{2^{n+1}}$ for the space of pure states of n qubits with respect to the trace norm [21], the identity of a state in the net can be transmitted using $O(2^n \log(1/\epsilon))$ bits of communication[1].

It is easy to see by a volume argument that the result of [21] is tight up to constant terms; that is, any ϵ-net must have size at least $(c/\epsilon)^{2^n}$ for some constant $c > 0$. But could we do better than this by using a protocol which is not based on simply discretising the space of quantum states via an ϵ-net? Questions of this nature are studied in the field of quantum communication complexity [11]. One of the first results in this area was work of Buhrman, Cleve and Wigderson [12] which implies that, in our terminology, Quantum State Sampling with $\epsilon = 0$ requires $\Omega(2^n)$ bits of classical communication. Their result is based on proving an $\Omega(2^n)$ lower bound on the deterministic communication complexity of a distributed version of the Deutsch-Jozsa problem in quantum query complexity. However, the complexity of this problem drops to $O(1)$ if ϵ is allowed to be nonzero.

Following this, a succession of results showed stronger separations between quantum and classical communication complexity. Raz [33] gave a communication task which could be solved by a two-way quantum protocol communicating $O(n)$ qubits, but which requires $\Omega(\sqrt{2^n})$ classical bits of classical communication. Bar-Yossef, Jayram and Kerenidis [5] showed that there is a communication task which can be solved by a *one-way* quantum protocol communicating $O(n)$ qubits, while any classical one-way bounded-error protocol must communicate $\Omega(\sqrt{2^n})$ bits. Gavinsky et al. [19] later proved a similar separation for a functional problem, i.e. one where Alice and Bob's task is to compute a function rather than sample from a distribution.

Finally, it was shown by Klartag and Regev that Quantum State Sampling requires $\Omega(\sqrt[3]{2^n})$ classical bits of communication between Alice and Bob [23], even if the communication is allowed to be two-way. This improved a previous result of Gavinsky [18], which implied that Quantum State Sampling requires $\Omega(\sqrt[8]{2^n}/\sqrt{n})$ bits of classical two-way communication. The lower bound of [23] is proven by considering a more restrictive problem known as the "vector in subspace" problem, which is defined as follows. Alice gets an n-qubit quantum state $|\psi\rangle$ and Bob gets a 2-outcome projective measurement $\{M, I - M\}$. Alice and Bob are promised that either $\langle\psi|M|\psi\rangle = 1$ or $\langle\psi|M|\psi\rangle = 0$; their task is to determine which is the case. This problem encompasses all one-way

[1] Note that the complexity achieved by this approach is better than would be obtained by simply writing down $|\psi\rangle$ in the computational basis, and truncating each amplitude after some number of digits. To achieve sufficient precision with this approach requires specifying each amplitude up to precision $O(\epsilon/\sqrt{2^n})$, giving an overall communication complexity of $O(2^n \log(2^n/\epsilon))$.

exact quantum protocols where Bob has two possible outputs [25].

The Quantum State Sampling problem has also been studied in the physics literature, where it is sometimes termed "classical teleportation". Toner and Bacon [36] showed that, in the case where Bob makes a projective measurement, Quantum State Sampling on one qubit can be solved exactly with 2 bits of communication from Alice to Bob (and shared randomness). An asymptotic protocol which encompasses POVMs and which uses slightly more communication was previously proposed by Cerf, Gisin and Massar [13]. Montina [29] gave an efficient classical protocol for the special case where Bob's measurement consists of two operators: the projector onto a pure state, and its complement (a similar result can be obtained from work of Kremer, Nisan and Ron [26]). Montina, Pfaffhauser and Wolf [30], and Montina and Wolf [31], related the asymptotic and one-shot communication complexities of exact classical simulation of general quantum channels to convex optimisation problems. The related problem of simulating the correlations obtained by measuring entangled states has also been studied by a number of authors; see [10, 11] for surveys.

Galvão and Hardy [17] considered a variant of the Quantum State Sampling problem, where Alice starts with a qubit in the state $|0\rangle$, the qubit is sent to Bob via a communication channel in which it undergoes a number of small rotations, and Bob is then asked to determine whether the final state of the qubit is $|0\rangle$ or $|1\rangle$, given that one of these is promised to be the case. Galvão and Hardy argued that any perfect classical simulation of this quantum protocol must use a system which can have infinitely many states, hence must transmit infinitely many bits of information. However, this lower bound does not hold for approximate simulations (e.g. the simulation method based on ϵ-nets mentioned above); in addition, it only holds when arbitrarily many operations can affect the qubit during its progress from Alice to Bob.

Thus, all these results leave open a natural question: could there exist a non-trivial classical protocol for Quantum State Sampling for fixed $\epsilon > 0$, i.e. one that transmits asymptotically less than 2^n bits? Such a protocol indeed exists for the vector in subspace problem: it was already shown by Raz [33] that this problem can be solved with bounded failure probability using $O(\sqrt{2^n})$ bits of classical one-way communication[2]. Could the same be true for the more general Quantum State Sampling problem? This question seems of interest on a fundamental, conceptual level: are quantum states "really" like an exponentially-long string of numbers, or do they have a more efficient representation?

1.1 Our results

Here we show that any classical communication protocol for Quantum State Sampling with sufficiently small constant inaccuracy $\epsilon > 0$ must transmit $\Omega(2^n)$ classical bits, even if the communication is allowed to be two-way. This immediately implies that any classical method for storing an arbitrary quantum state such that measurements can be approximately simulated on that state must store $\Omega(2^n)$ classical bits. This can be seen as an "anti-Holevo" theorem: a quantum state of n qubits can only store at most n bits [22], but $\Omega(2^n)$ classical bits are required to store n qubits.

This result is based on proving a quantum-classical separation for the following special case of Quantum State Sampling, which we call Distributed Fourier Sampling:

- Alice is given a function $f : \{0,1\}^n \to \{\pm 1\}$.
- Bob is given a function $g : \{0,1\}^n \to \{\pm 1\}$.

[2] This result was stated in [33] but the proof has not appeared. We include a proof in Appendix A.

- Their task is for one party (say Bob) to approximately sample from the distribution p_{fg} on n-bit strings s where

$$p_{fg}(s) = \left(\frac{1}{2^n} \sum_{x \in \{0,1\}^n} (-1)^{s \cdot x} f(x) g(x) \right)^2 \tag{1}$$

  and $s \cdot x = \sum_{i=1}^n s_i x_i$. That is, Bob must output a sample from any distribution $\widetilde{p}_{fg}$ such that $\|\widetilde{p}_{fg} - p_{fg}\|_1 \leq \epsilon$, for some constant ϵ.

The title "Distributed Fourier Sampling" refers to the fact that the distribution which must be sampled from is the square of the Fourier transform of the function $fg(x) = f(x)g(x)$ over $\mathbb{Z}_2^n$.

For conciseness, we henceforth write $N = 2^n$. Distributed Fourier Sampling can be solved with n qubits of one-way communication and $\epsilon = 0$. Alice constructs the state $|\psi_f\rangle = \frac{1}{\sqrt{N}} \sum_{x \in \{0,1\}^n} f(x)|x\rangle$ and sends it to Bob. Bob then applies the unitary operator defined by $U_g|x\rangle = g(x)|x\rangle$ to $|\psi_f\rangle$ to produce $|\psi_{fg}\rangle = \frac{1}{\sqrt{N}} \sum_{x \in \{0,1\}^n} f(x)g(x)|x\rangle$. Finally, Bob applies a Hadamard gate to each qubit of $|\psi_{fg}\rangle$ and measures in the computational basis. The resulting distribution is exactly p_{fg}.

By contrast, we have the following result:

Theorem 1. *For sufficiently small constant $\epsilon > 0$, any classical two-way communication protocol with shared randomness for Distributed Fourier Sampling must communicate $\Omega(N)$ bits.*

1.2 Consequences of the lower bound

Theorem 1 immediately implies similar separations in related models.

Query complexity of sampling problems. We can use Distributed Fourier Sampling to obtain a lower bound in the query model on the classical complexity of the (non-distributed) Fourier Sampling problem [8, 2]. In this problem, we are given the ability to query (evaluate) an unknown function $h : \{0,1\}^n \to \{\pm 1\}$, which corresponds to an input of size N. Our task is to approximately sample from the Fourier spectrum of h, i.e. the distribution p_h on bit-strings $s \in \{0,1\}^n$ where

$$p_h(s) = \left(\frac{1}{2^n} \sum_{x \in \{0,1\}^n} (-1)^{s \cdot x} h(x) \right)^2 .$$

More precisely, we are asked to output a sample from any distribution $\widetilde{p}_h$ such that $\|\widetilde{p}_h - p_h\|_1 \leq \epsilon$. This problem can be solved exactly with 1 quantum query to h by constructing the state $\frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} h(x)|x\rangle$, applying a Hadamard gate to each qubit, and measuring in the computational basis.

Any randomised classical protocol solving Fourier Sampling using t queries immediately implies a classical two-way communication protocol with shared randomness for Distributed Fourier Sampling communicating at most $2t$ bits, via a standard reduction [27]. Alice and Bob simulate the procedure for Fourier Sampling, and whenever they want to query $h(x)$, they replace the query with evaluating $f(x)g(x)$ using 2 bits of communication. Therefore, Theorem 1 implies a corresponding lower bound on the query complexity of Fourier Sampling:

Corollary 2. *For sufficiently small constant $\epsilon > 0$, any randomised classical algorithm solving Fourier Sampling on N input bits must make $\Omega(N)$ queries.*

A lower bound of $\Omega(N/\log N)$ queries on Fourier Sampling was previously shown by Aaronson and Ambainis [2], who conjectured that this bound was tight, not only for Fourier Sampling, but for *all* sampling problems that can be solved with 1 quantum query. Corollary 2 refutes this conjecture. Interestingly, it was also shown by the same authors that any partial boolean function which can be computed by a bounded-error quantum algorithm making $t = O(1)$ queries can be computed by a bounded-error classical algorithm making $O(N^{1-1/(2t)})$ queries [2]. Thus, to see maximal quantum-classical query separations, we are required to go beyond computing boolean functions.

An alternative, direct proof of Corollary 2 is presented in Appendix B. Following the completion of this work, I learned that Scott Aaronson and Lijie Chen have recently obtained an independent proof of this result [3].

Nonlocality problems. Using a standard mapping between communication protocols and entanglement [11], we can obtain a distribution $\mathcal{D}$ which can be sampled from exactly with no communication between the parties if Alice and Bob share n Bell pairs, but such that any classical procedure for sampling from $\mathcal{D}$ up to distance ϵ in ℓ_1 norm, for some constant $\epsilon > 0$, requires $\Omega(2^n)$ bits of classical communication. A similar lower bound for *exact* sampling from the same distribution $\mathcal{D}$ was previously shown by Brassard, Cleve and Tapp [9], but the bounded-error case was called "an important open question" by Toner and Bacon [36].

To obtain $\mathcal{D}$, we define a variant of the Distributed Fourier Sampling problem. Alice and Bob are again each given a function, f and g respectively. However, this time they are asked to sample from a distribution on pairs of bit-strings $s, t \in \{0,1\}^n$ (where Alice outputs s, Bob outputs t) of the following form, up to ℓ_1 distance ϵ: a distribution where $\Pr[s \oplus t = u] = p_{fg}(u)$ for all $u \in \{0,1\}^n$, where p_{fg} is defined as in (1). We call this problem Doubly Distributed Fourier Sampling (DDFS).

The quantum protocol for DDFS proceeds as follows. Alice and Bob share a maximally entangled state $\frac{1}{\sqrt{N}}\sum_{x\in\{0,1\}^n}|x\rangle|x\rangle$. Alice and Bob each apply the unitary operators U_f and U_g, defined by $U_f|x\rangle = f(x)|x\rangle$, $U_g|x\rangle = g(x)|x\rangle$, to their half of the state to produce the state $|\phi_{fg}\rangle = \frac{1}{\sqrt{N}}\sum_{x\in\{0,1\}^n} f(x)g(x)|x\rangle|x\rangle$. They each then apply Hadamard gates to each qubit of their half of $|\phi_{fg}\rangle$, measure in the computational basis, and output the result.

The final state produced before measuring is

$$\frac{1}{N^{3/2}}\sum_{x\in\{0,1\}^n} f(x)g(x)\left(\sum_{s\in\{0,1\}^n}(-1)^{s\cdot x}|s\rangle\right)\left(\sum_{t\in\{0,1\}^n}(-1)^{t\cdot x}|t\rangle\right),$$

so for each pair of bit-strings (s,t) such that $s \oplus t = u$, the probability that Alice and Bob see that pair of bit-strings is

$$\frac{1}{N^3}\left(\sum_{x\in\{0,1\}^n}(-1)^{x\cdot(s\oplus t)}f(x)g(x)\right)^2 = \frac{p_{fg}(u)}{N}.$$

Therefore, the probability that they see some pair of bit-strings (s,t) such that $s\oplus t = u$ is exactly $p_{fg}(u)$. On the other hand, any classical protocol for DDFS approximating the output distribution up to ℓ_1 accuracy ϵ and communicating c bits gives a protocol for Distributed Fourier Sampling up to accuracy ϵ which communicates $c+n$ bits. This protocol proceeds as follows: after receiving outcomes (s,t) from their DDFS protocol, Alice sends s to Bob, who outputs $s \oplus t$. If the probability that Alice and Bob output (s,t) was $\widetilde{q}_{fg}(s,t)$, then as the DDFS protocol achieved ℓ_1 distance at most ϵ from a distribution where $\Pr[s\oplus t = u] = p_{fg}(u)$ for all $u \in \{0,1\}^n$, we have

$$\sum_{s,t\in\{0,1\}^n}|q_{fg}(s,t) - \widetilde{q}_{fg}(s,t)| \le \epsilon$$

for some distribution q_{fg} such that for all $u \in \{0,1\}^n$, $\sum_{s \oplus t = u} q_{fg}(s,t) = p_{fg}(u)$. So the ℓ_1 distance between the output distribution and the desired distribution is

$$\begin{aligned}
\sum_{u \in \{0,1\}^n} |p_{fg}(u) - \sum_{s \oplus t = u} \widetilde{q}_{fg}(s,t)| &= \sum_{u \in \{0,1\}^n} \left| \sum_{s \oplus t = u} \frac{q_{fg}(s,t)}{p_{fg}(u)} p_{fg}(u) - \widetilde{q}_{fg}(s,t) \right| \\
&\leq \sum_{u \in \{0,1\}^n} \sum_{s \oplus t = u} |q_{fg}(s,t) - \widetilde{q}_{fg}(s,t)| \\
&= \sum_{s,t \in \{0,1\}^n} |q_{fg}(s,t) - \widetilde{q}_{fg}(s,t)| \leq \epsilon.
\end{aligned}$$

We have obtained the following corollary:

Corollary 3. *For sufficiently small constant $\epsilon > 0$, Doubly Distributed Fourier Sampling requires $\Omega(N)$ bits of classical two-way communication.*

This separation is tight up to constant factors, because the distribution obtained by measuring an entangled state of local dimension N can be simulated up to constant accuracy using $O(N)$ bits of communication, via a similar ϵ-net construction as for Quantum State Sampling.

Finally, we remark that the results given here on Fourier Sampling and its distributed variant are connected to some of the earliest works on quantum computation: the first exponential separation between exact quantum and classical algorithms via the Deutsch-Jozsa algorithm [16], the first super-polynomial separation between quantum and randomised classical algorithms via recursive Fourier sampling [8], and the first exponential separations between exact quantum and classical communication complexity via the distributed version of the Deutsch-Jozsa algorithm [12, 9]. It is remarkable that, around 20 years after these pioneering results, Fourier Sampling continues to be a rich vein from which to mine quantum-classical separations.

1.3 Sketch of the proof of the main result

In the remainder of the paper, we prove the claimed lower bound on the classical communication complexity of Distributed Fourier Sampling. We first give an outline of the proof, which combines the following ingredients:

1. If there is a classical protocol for approximately sampling from p_{fg} up to constant inaccuracy ϵ, there is a classical protocol with two outcomes (accept or reject) which communicates the same number of bits and accepts with probability very close to

$$\left(\frac{\langle f,g\rangle}{N}\right)^2 := \left(\frac{1}{N} \sum_{x \in \{0,1\}^n} f(x)g(x)\right)^2; \qquad (2)$$

   more specifically, a protocol which accepts with probability $\widetilde{p}(f,g)$ such that the average of $|\widetilde{p}(f,g) - (\langle f,g\rangle/N)^2|$ over f and g is at most ϵ/N. A similar idea was used in [2] in the setting of query complexity.

2. For sufficiently small $\epsilon > 0$, any classical protocol whose acceptance probability is this close to $(\langle f,g\rangle/N)^2$ on average over f and g must communicate $\Omega(N)$ bits. This is the main technical challenge. To address it, we extend a remarkably powerful proof technique used by

Chakrabarti, Kondapally and Wang [14] to show a lower bound on the information complexity of the gap-orthogonality problem (qv). This bound in turn extended ideas from previous proofs of Chakrabarti and Regev [15] and Sherstov [34] of an $\Omega(N)$ lower bound on the communication complexity of the gap-Hamming problem (a promise version of the problem of determining Hamming distance).

The proof is based on the following steps:

(a) Assume that there exists a protocol which accepts with probability $\widetilde{p}(f,g)$ such that $\mathbb{E}_{f,g}[|\widetilde{p}(f,g) - (\langle f,g\rangle/N)^2|] = O(1/N)$, and which communicates at most γN bits for some small constant $\gamma > 0$. First show that this corresponds to a partition of Alice and Bob's set of inputs, $\{\pm 1\}^N \times \{\pm 1\}^N$, into rectangles (subsets of inputs of the form $R = X \times Y$) such that many rectangles are large, i.e. $|X|, |Y| \geq 2^{\gamma' N}$ for some $\gamma' > 0$. (This part is standard.) Here a rectangle R represents the final state of a deterministic communication protocol.

(b) Think of f and g as strings of ± 1's of length N each, and relabel them x and y. Show that, for all large enough rectangles $R = X \times Y$, and all choices of $\nu_{R,y} \in \mathbb{R}$, we have $\mathbb{E}_{x \in X, y \in Y}[|\nu_{R,y} - (\langle x,y\rangle/N)^2|] = \Omega(1/N)$. $\nu_{R,y}$ is Bob's probability of acceptance, based on the communication transcript of the two parties (represented by the rectangle R) and his input y. This step is a proof by contradiction which in turn can be broken into two parts, following the approach of [14]:

    i. For any large enough subsets of inputs $X, Y \subseteq \{\pm 1\}^N$, there exists $y \in Y$ such that the distribution of inner products $\langle x,y\rangle$ over randomly chosen inputs $x \in X$ is roughly the same as a scaled and shifted Gaussian, so $|\langle x,y\rangle|$ is not too tightly concentrated around any value. Note that the distribution is clearly not always peaked around 0 like a standard Gaussian; for example, consider sets X and Y consisting of strings whose first $N/2$ entries are all the same.

    ii. For this particular y, if $\mathbb{E}_{x \in X}[|\nu_{R,y} - (\langle x,y\rangle/N)^2|] \ll 1/N$, then $|\langle x,y\rangle|$ must be tightly concentrated around some value.

    Combining these two parts gives the desired contradiction. As in [14] (and also [15]), one important technicality is that we need to deal with the "Gaussian version" of the random variable $\langle x,y\rangle$ for some fixed y, obtained by replacing x with the random variable $\widetilde{x}$ given by multiplying each entry of x by a normally distributed random variable. Although the flow of the proof is the same as [14], we encounter some additional difficulties. For the first part, we need to show that $\langle \widetilde{x}, y\rangle$ is not too concentrated around any value (as opposed to just "not too large" as in [14]). For the second part, we similarly need to handle the case where $|\langle x,y\rangle|$ is concentrated around some nonzero (and potentially large) value separately.

Combining parts (a) and (b) proves the desired lower bound: if many rectangles are large, and Bob does not sample accurately when the final state of the protocol corresponds to a large rectangle, Bob must not sample accurately on most inputs.

It is instructive to check why we cannot just use existing communication complexity results to prove a lower bound on DFS. First, by the $O(\sqrt{2^n})$ classical upper bound of Raz [33] for any partial boolean function, we can see that we need to look beyond protocols whose acceptance and rejection probabilities are separated by an additive constant. Looking at (2), one might naturally guess that a tight lower bound could be proven by considering inputs obeying the promise that (for example) either $\langle f,g\rangle = 0$, or $|\langle f,g\rangle| \geq 3\epsilon\sqrt{N}$. In the former case, the protocol should accept with probability at most ϵ/N, and in the latter case the protocol should accept with probability at least $(3\epsilon - \epsilon)/N = 2\epsilon/N$. So there is a multiplicative constant separating the acceptance probabilities in the two cases.

This is a variant of the well-studied *gap-orthogonality* problem [34, 14]. In this problem (in our notation), Alice and Bob are asked to accept if $|\langle f, g\rangle| \geq a\sqrt{N}$, and reject if $|\langle f, g\rangle| \leq b\sqrt{N}$, for some constants $a > b$. For suitably chosen values of a and b, an $\Omega(N)$ lower bound is known on the communication complexity of the gap-orthogonality problem if Alice and Bob are asked to succeed in this task with probability at least $2/3$ [15, 37, 34]. Thus, extending this result to hold in the nonstandard setting where acceptance and rejection probabilities are small and separated by a multiplicative constant would suffice to prove our desired result.

Further, Göös and Watson [20] showed that communication lower bounds of this form can be proven using the corruption method [38, 6, 24], an important lower bound technique in communication complexity, used in particular by Sherstov to prove his $\Omega(N)$ lower bound for gap-orthogonality [34]. If we had a corruption bound of $\Omega(N)$ for the gap-orthogonality problem, this would imply the result we need.

But in fact such a bound cannot exist, because Göös and Watson also showed that the corruption bound is *equivalent* to the communication complexity of an optimal protocol whose acceptance and rejection probabilities can be arbitrarily small, but are separated by a multiplicative constant [20]. And there is indeed a nontrivial protocol of this form for gap-orthogonality: query $\sqrt{N}$ random bits of Alice and Bob's inputs, and accept if they are all different, or all the same. If Alice and Bob's strings differ at a $1/2 + \Delta/\sqrt{N}$ fraction of positions, the probability of acceptance is precisely

$$\left(\frac{1}{2} + \frac{\Delta}{\sqrt{N}}\right)^{\sqrt{N}} + \left(\frac{1}{2} - \frac{\Delta}{\sqrt{N}}\right)^{\sqrt{N}} = \frac{1}{2^{\sqrt{N}}}\left(\left(1 + \frac{2\Delta}{\sqrt{N}}\right)^{\sqrt{N}} + \left(1 - \frac{2\Delta}{\sqrt{N}}\right)^{\sqrt{N}}\right).$$

This is roughly equal to $2^{-\sqrt{N}}(e^{2\Delta} + e^{-2\Delta}) = 2^{1-\sqrt{N}}\cosh(2\Delta)$. For Δ values separated by a constant factor, the acceptance probabilities will also be separated by a constant factor (which can be made arbitrarily large by taking the AND of multiple runs).

It therefore does not seem possible to use gap-orthogonality to prove the desired lower bound via standard techniques. So how did Sherstov prove a corruption bound of $\Omega(N)$ for gap-orthogonality? He considered the negation of the problem we consider here (i.e. reversing the roles of acceptance and rejection), which *does* have such a lower bound. However, it is not clear how to use this problem to prove a bound on the original Fourier sampling task, as the acceptance probabilities of the quantum protocol do not correspond directly to the negated problem.

We finally remark that, using a connection between the corruption bound and Bell inequalities, the gap-orthogonality problem was also recently used by Laplante et al. [28] to give inefficiency-resistant Bell inequalities with large violations.

We now describe the proof of Theorem 1 in detail.

2 The lower bound on Distributed Fourier Sampling

Assume towards a contradiction that there is a classical two-way communication protocol where Alice and Bob have access to shared randomness, communicate c bits, and such that at the end of the protocol Bob outputs a sample from a distribution $\widetilde{p}_{fg}$ such that $\sum_{s\in\{0,1\}^n} |\widetilde{p}_{fg}(s) - p_{fg}(s)| \leq \epsilon$.

(Recall that $p_{fg}(s) = [2^{-n} \sum_{x \in \{0,1\}^n} (-1)^{s \cdot x} f(x) g(x)]^2$.) Then in particular

$$\mathbb{E}_{f,g}\left[\sum_{s \in \{0,1\}^n} |\widetilde{p}_{fg}(s) - p_{fg}(s)|\right] \leq \epsilon,$$

where f and g are picked from the uniform distribution. Interchanging the expectation and the sum, there must exist some s such that $\mathbb{E}_{f,g}[|\widetilde{p}_{fg}(s) - p_{fg}(s)|] \leq \epsilon/N$. We now observe that we can assume $s = 0^n$ without loss of generality. If not, then consider the protocol where Alice replaces f with the function $f'(x) = (-1)^{s \cdot x} f(x)$. Then $p_{f'g}(0^n) = p_{fg}(s)$, and if the pair (f, g) was uniformly distributed, so is the pair (f', g). Associating the output 0^n with acceptance, and any other outcome with rejection, we obtain a protocol Π that communicates c bits and accepts with probability $q(f, g)$, such that

$$\mathbb{E}_{f,g}\left[\left|q(f,g) - \left(\frac{\langle f, g\rangle}{N}\right)^2\right|\right] \leq \frac{\epsilon}{N}.$$

So the output of Π is correlated well with $(\langle f, g\rangle/N)^2$.

It will be convenient for the rest of the proof to think of f and g as corresponding to strings $x, y \in \{\pm 1\}^N$, to use the notation $x[i]$ for the i'th entry of x, and to reserve the notation x_i for elements of a sequence of strings $x_1, x_2, \dots$. In this notation, then, we assume that we have a classical protocol which communicates c bits, such that Bob accepts with probability $q(x, y)$ on input (x, y), where $x, y \in \{\pm 1\}^N$, and

$$\mathbb{E}_{x,y}\left[\left|q(x,y) - \left(\frac{\langle x, y\rangle}{N}\right)^2\right|\right] \leq \frac{\epsilon}{N}. \tag{3}$$

Here $\langle x, y\rangle = \sum_i x[i]y[i]$ and the expectation is with respect to the uniform distribution μ on pairs (x, y). We now state the key technical claim that we will prove. For $X \subseteq \{\pm 1\}^N$, let $\mu(X)$ denote the measure of X under the uniform distribution, i.e. $\mu(X) = |X|/2^N$. A rectangle is a subset $R \subseteq \{\pm 1\}^N \times \{\pm 1\}^N$ of the form $R = X \times Y$ for $X, Y \subseteq \{\pm 1\}^N$. Then we will show:

Lemma 4. *There exist constants $\gamma, \delta > 0$ such that, for sufficiently large N, all rectangles $R = X \times Y$ such that $\mu(X), \mu(Y) > 2^{-\gamma N}$, and all sequences ν_y, we have $\mathbb{E}_{x \in X, y \in Y}[|\nu_y - (\langle x, y\rangle/N)^2|] \geq \delta/N$.*

We first show that Lemma 4, together with the above assumption about the existence of an efficient classical protocol, implies the desired contradiction. As the inputs (x, y) are random, we can assume without loss of generality that the classical protocol is deterministic [27], followed by Bob accepting with probability $q(x, y)$. Any such protocol partitions the $2^N \times 2^N$ matrix $M_{xy} = (\langle x, y\rangle/N)^2$ into at most 2^c rectangles R such that, at the end of the protocol, Alice and Bob know that their inputs are picked from R. Then Bob's acceptance probability depends only on R and his input y. By Lemma 4, there exist $\gamma, \delta > 0$ such that, for all rectangles $R = X \times Y$ such that $\mu(X), \mu(Y) > 2^{-\gamma N}$, and all choices of $\nu_{R,y}$, we have $\mathbb{E}_{x \in X, y \in Y}[|\nu_{R,y} - (\langle x, y\rangle/N)^2|] \geq \delta/N$.

Call a rectangle $R = X \times Y$ such that $\mu(X), \mu(Y) > 2^{-\gamma N}$ large, and otherwise small. Then

$$\begin{aligned}
\mathbb{E}_{(x,y)\sim\mu}\left[\left|q(x,y) - \left(\frac{\langle x,y\rangle}{N}\right)^2\right|\right] &= \sum_{x,y} \mu(x,y)\left|q(x,y) - \left(\frac{\langle x,y\rangle}{N}\right)^2\right| \\
&= \sum_{R}\sum_{(x,y)\in R} \mu(x,y)\left|q(x,y) - \left(\frac{\langle x,y\rangle}{N}\right)^2\right| \\
&= \sum_{R}\frac{\mu(R)}{|R|}\sum_{(x,y)\in R}\left|\nu_{R,y} - \left(\frac{\langle x,y\rangle}{N}\right)^2\right| \\
&\geq \sum_{\text{large } R} \mu(R)\frac{\delta}{N}.
\end{aligned}$$

Next assume that $c \leq \gamma N - 1$. Then

$$\sum_{\text{large } R} \mu(R) = 1 - \sum_{\text{small } R} \mu(R) \geq 1 - 2^{c-\gamma N} \geq \frac{1}{2}.$$

So we have $\mathbb{E}_{(x,y)\sim\mu}[|q(x,y) - (\langle x,y\rangle/N)^2|] \geq \frac{\delta}{2N}$, contradicting (3) for $\epsilon < \delta/2$.

It remains to prove Lemma 4. This is a generalisation of a result by Sherstov [34, Theorem 3.3] (see [15, 37] for related previous work), who showed that there exist constants $\gamma, \epsilon > 0$ such that, for all rectangles $R = X \times Y$ such that $\mu(X), \mu(Y) > 2^{-\gamma N}$, we have $\Pr_{x\in X, y\in Y}[|\langle x,y\rangle| \geq \sqrt{N}/4] \geq \epsilon$. This implies $\mathbb{E}_{x\in X, y\in Y}[(\langle x,y\rangle/N)^2] \geq \delta/N$ for some $\delta > 0$, which would be the bound we need for $\nu_y = 0$. Here we need to generalise to ν_y being nonzero and potentially depending on y. We achieve this using proof techniques of Chakrabarti, Kondapally and Wang [14], who generalised Sherstov's result to prove a lower bound of the form $\Pr_{x\in X, y\in Y}[|\langle x,y\rangle| \geq b\sqrt{N}] \geq \epsilon$ for large b. This required substantial additional technical work, building on ideas from the previous paper of Chakrabarti and Regev [15]. Here we generalise the result of [14] further by, roughly speaking, proving anticoncentration of $|\langle x,y\rangle|$ around an arbitrary point ν_y, rather than around 0; the overall flow of the proof is the same.

First we state some results from [14] which we will need.

Fact 5 (Chakrabarti, Kondapally and Wang [14], generalising Sherstov [34])**.** *There exists $\zeta_0 > 0$ such that, for all sufficiently large N and all ζ satisfying $0 < \zeta \leq \zeta_0$, the following holds: Set $k = \lceil\sqrt{\zeta}N\rceil$ and assume that $Y \subseteq \{\pm 1\}^N$ satisfies $|Y| \geq 2^{N-\zeta N-1}$. Then there exist $y_1, \dots, y_k \in Y$ such that for all $j \in \{1, \dots, k\}$, we have*

$$\|\operatorname{proj}_{\operatorname{span}\{y_1,\dots,y_{j-1}\}} y_j\| \leq 2\zeta^{1/4}\sqrt{N}.$$

We henceforth use the notation $y_1, \dots, y_k$ for the set of nearly-orthogonal vectors within some set Y guaranteed to exist by Fact 5. Let $\widetilde{y}_j$ denote the exactly orthogonal vectors obtained from $y_1, \dots, y_k$ via the Gram-Schmidt procedure, and let y_j^* denote the normalisations $y_j^* = y_j/\|y_j\|$.

We think of x as a discrete random variable picked uniformly from X and define a continuous random variable $\widetilde{x} \in \mathbb{R}^N$ based on x by setting $\widetilde{x}[j] = |w_j|\, x[j]$ for all $j \in \{1, \dots, n\}$, where $w_j \sim N(0,1)$ is distributed according to the standard normal distribution. Note that $\mathbb{E}_{w_1,\dots,w_N}[\widetilde{x}] = \sqrt{2/\pi}x$. Next we define a sequence of random variables q_j by $q_j = \langle \widetilde{x}, y_j\rangle/\sqrt{N}$. If Bob's input is y_j, $\langle x, y_j\rangle/\sqrt{N}$ is the (suitably scaled) random variable controlling how inaccurate Bob's sample will be – if the random variable is highly concentrated, then Bob's sampling will be accurate. Then q_j is the random variable obtained from this by replacing x with its continuous variant $\widetilde{x}$.

For a continuous probability distribution P define $D_N(P) := D(P\|N(0,1))$, where $D(\cdot\|\cdot)$ is the relative entropy,

$$D(P\|Q) := \int P(x)\ln(P(x)/Q(x))dx.$$

Pinsker's inequality states that, for any P and Q, $\|P-Q\|_1 \leq \sqrt{2D(P\|Q)}$. Also define $D(X \mid Y) := \mathbb{E}_Y[D_N(X \mid Y)]$, i.e. $D(X \mid Y)$ is the expected "distance" over Y of the conditional probability distribution $X \mid Y$ from $N(0,1)$. We can now state the next result from [14] that we need:

Lemma 6 (Encapsulates claims of Chakrabarti, Kondapally and Wang [14, Section 5])**.** *Assume that the preconditions of Fact 5 hold. Then we can write*

$$q_j = r_j z_j + s_j, \text{ where } \quad r_j = \frac{\langle y_j, y_j^* \rangle}{\sqrt{N}}, \quad z_j = \langle \widetilde{x}, y_j^* \rangle, \quad s_j = \frac{1}{\sqrt{N}} \sum_{i=1}^{j-1} \langle y_j, y_i^* \rangle \langle \widetilde{x}, y_i^* \rangle.$$

For all $j \in \{1, \ldots, k\}$, we have $1 - 4\sqrt{\zeta} \leq r_j \leq 1$, where ζ is the constant from Fact 5. In addition, there exists $j \in \{1, \ldots, k\}$ such that $D(z_j \mid s_j) \leq \sqrt{\zeta}$.

Following [15], let tail(x) denote the tail of the standard normal distribution:

$$\text{tail}(x) := \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-x^2/2} dx = \frac{1}{2}\,\text{erfc}\left(\frac{x}{\sqrt{2}}\right).$$

We are finally ready to prove Lemma 4.

Lemma 4 (restated). *There exist constants $\gamma, \delta > 0$ such that, for sufficiently large N, all rectangles $R = X \times Y$ such that $\mu(X), \mu(Y) > 2^{-\gamma N}$, and all sequences ν_y, we have $\mathbb{E}_{x \in X, y \in Y}[|\nu_y - (\langle x, y \rangle/N)^2|] \geq \delta/N$.*

Proof. Assume towards a contradiction the negation of the statement of the lemma, namely that for some small constants $\gamma, \delta > 0$ there exist a rectangle $R = X \times Y$ and a sequence ν_y such that $\mu(X), \mu(Y) > 2^{-\gamma N}$ and $\mathbb{E}_{x \in X, y \in Y}[|\nu_y - (\langle x, y \rangle/N)^2|] \leq \delta/N$. Set

$$Y' := \left\{ y \in Y : \mathbb{E}_{x \in X}\left[\left|\nu_y - \left(\frac{\langle x, y \rangle}{N}\right)^2\right|\right] \leq \frac{2\delta}{N} \right\}. \tag{4}$$

We have $|Y'| \geq |Y|/2 > 2^{(1-\gamma)N+1}$ by Markov's inequality. Y' is the set of "good" inputs y such that, on average over x, Bob's acceptance probability ν_y is close to the desired value $(\langle x, y \rangle/N)^2$. For small enough $\gamma > 0$, we can apply Fact 5 to Y' to obtain nearly-orthogonal vectors y_j and corresponding random variables q_j, r_j, z_j, s_j as in the previous discussion.

By Lemma 6, there exists j such that $D(z_j \mid s_j) \leq \sqrt{\zeta}$, where ζ is a small constant, $z_j = \langle \widetilde{x}, y_j^* \rangle$, and s_j is some other random variable. Henceforth fix this j and write $r := r_j$, $z := z_j$, $s := s_j$, $q := q_j = \langle \widetilde{x}, y_j \rangle / \sqrt{N}$. We now prove concentration and anticoncentration bounds on q, in terms of upper and lower bounds (respectively) on $\Pr[|q - a| \geq b]$ for some $a, b \in \mathbb{R}$. We will show that the lower bound is higher than the upper bound, giving our desired contradiction.

Anticoncentration bound. This part will be based on Lemma 6, i.e. that q can (roughly speaking) be thought of as being normally distributed. Define the set

$$\text{Good} = \{s' \in \mathbb{R} : D(z \mid s = s') \leq \zeta^{1/4}\}.$$

Then by Markov's inequality $\Pr[s \notin \text{Good}] \le \zeta^{1/4}$. Fix $a \in \mathbb{R}$. Either $\Pr[s \ge a \mid s \in \text{Good}] \ge 1/2$ or $\Pr[s \le a \mid s \in \text{Good}] \ge 1/2$. First assume the former. Then $D(z \mid s \ge a \wedge s \in \text{Good}) \le 2\zeta^{1/4}$. By Pinsker's inequality, the total variation distance between $N(0,1)$ and the distribution of z conditioned on $s \le a$ and $s \in \text{Good}$ is at most $2\zeta^{1/8}$. So

$$\begin{aligned}
\Pr[q \ge a+b] &= \Pr[rz + s \ge a+b] \\
&= \Pr[rz+s \ge a+b \mid s \ge a \wedge s \in \text{Good}] \Pr[s \ge a \mid s \in \text{Good}] \Pr[s \in \text{Good}] \\
&\ge \frac{1}{2}(1-\zeta^{1/4}) \Pr[rz+s \ge a+b \mid s \ge a \wedge s \in \text{Good}] \\
&\ge \frac{1}{2}(1-\zeta^{1/4}) \Pr[z \ge b/r \mid s \ge a \wedge s \in \text{Good}] \\
&\ge \frac{1}{2}(1-\zeta^{1/4})(\text{tail}(b/r) - 2\zeta^{1/8}) \\
&\ge \frac{1}{2}(1-\zeta^{1/4}) \left(\text{tail}\left(\frac{b}{1-4\sqrt{\zeta}}\right) - 2\zeta^{1/8}\right).
\end{aligned}$$

If instead $\Pr[s \le a \mid s \in \text{Good}] \ge 1/2$, we follow the same argument, but starting with $\Pr[q \le a-b]$ and adjusting appropriately throughout. As one or other of these two cases must hold, the final result is the bound

$$\Pr[|q-a| \ge b] \ge \frac{1}{2}(1-\zeta^{1/4}) \left(\text{tail}\left(\frac{b}{1-4\sqrt{\zeta}}\right) - 2\zeta^{1/8}\right)$$

valid for any $a, b \in \mathbb{R}$.

Concentration bound. This part uses the assumption that Bob accepts with probability close to $(\langle x,y\rangle/N)^2$. Write $\nu := \nu_{y_j}$, and assume that $y_j = 1^N$ such that $\langle x,y\rangle = \sum_i x[i]$. This is without loss of generality as otherwise we can flip the signs of entries of each x without affecting $\langle x,y\rangle$. Let t be the random variable $t = \frac{1}{\sqrt{N}} \sum_i x[i]$.

It was shown in [15] that as $N \to \infty$, q converges in distribution to $\sigma v + \sigma' t$, where $\sigma = \sqrt{1-2/\pi}$, $\sigma' = \sqrt{2/\pi}$ and $v \sim N(0,1)$ is independent of t. That is, the discrete and continuous parts of q can be handled separately. By (4) we have that $\mathbb{E}_{x \in X}\left[\left|\nu - \left(\frac{\langle x,y\rangle}{N}\right)^2\right|\right] \le \frac{2\delta}{N}$, so in terms of t we have

$$\mathbb{E}_t\left[\left|\nu - \frac{t^2}{N}\right|\right] \le \frac{2\delta}{N}.$$

We now split into two cases. First assume that $\nu \le \delta/N$. Then

$$\mathbb{E}_t\left[\left|\nu - \frac{t^2}{N}\right|\right] \ge \mathbb{E}_t\left[\frac{t^2}{N}\right] - \nu \ge \mathbb{E}_t\left[\frac{t^2}{N}\right] - \frac{\delta}{N},$$

so $\mathbb{E}_t[t^2] \le 3\delta$ and by Markov's inequality,

$$\Pr_t\left[|t| \ge \sqrt{\frac{3\delta}{c}}\right] \le c$$

for any $c > 0$. On the other hand, if $\nu \ge \delta/N$, then

$$\mathbb{E}_t\left[\left|\nu - \frac{t^2}{N}\right|\right] = \mathbb{E}_t\left[\left|\sqrt{\nu} - \frac{|t|}{\sqrt{N}}\right|\left|\sqrt{\nu} + \frac{|t|}{\sqrt{N}}\right|\right] \ge \mathbb{E}_t\left[\left|\sqrt{\nu} - \frac{|t|}{\sqrt{N}}\right|\right]\sqrt{\frac{\delta}{N}},$$

so $\mathbb{E}_t\left[\left|\sqrt{N\nu}-|t|\right|\right] \leq 2\sqrt{\delta}$ and, using Markov's inequality again,

$$\Pr_t\left[\left|\sqrt{N\nu}-|t|\right| \geq \frac{2\sqrt{\delta}}{c}\right] \leq c$$

for any $c > 0$.

By convergence of q to $\sigma v + \sigma' t$, for sufficiently large N we have

$$\Pr[||q|-a| \geq b] \leq 2\Pr[||\sigma v + \sigma' t| - a| \geq b]$$

for any a, b. We now estimate

$$\begin{aligned}\Pr[||q|-a| \geq b] &\leq 2\Pr[||\sigma v + \sigma' t| - a| \geq b] \\ &\leq 2(\Pr[||\sigma v + \sigma' t| - a| \geq b \mid |v| \leq b'] + \Pr[|v| \geq b']) \\ &\leq 2\Pr[||\sigma' t| - a| \geq b - \sigma b'] + 4\operatorname{tail}(b')\end{aligned}$$

for any $b' \geq 0$, so

$$\Pr[||q|-a| \geq b] \leq 2\Pr\left[||t|-a| \geq \frac{b-\sigma b'}{\sigma'}\right] + 4\operatorname{tail}(b').$$

If we were in the case $\nu \leq \delta/N$ above, take $a = 0$, $b' = (b - \sigma'\sqrt{3\delta/c})/\sigma$, $c = \operatorname{tail}(2b)$ to obtain the bound

$$\Pr[||q|-a| \geq b] \leq 2c + 4\operatorname{tail}(b') = 2\operatorname{tail}(2b) + 4\operatorname{tail}\left(\frac{b - \sigma'\sqrt{3\delta/\operatorname{tail}(2b)}}{\sigma}\right).$$

If we were in the case $\nu \geq \delta/N$ above, take $a = \sqrt{N\nu}$, $b' = (b - 2\sigma'\sqrt{\delta}/c)/\sigma$, $c = \operatorname{tail}(2b)$ to obtain

$$\Pr[||q|-a| \geq b] \leq 2c + 4\operatorname{tail}(b') = 2\operatorname{tail}(2b) + 4\operatorname{tail}\left(\frac{b - 2\sigma'\sqrt{\delta}/\operatorname{tail}(2b)}{\sigma}\right).$$

In either case, for sufficiently small $\delta > 0$, $\Pr[||q|-a| \geq b] \leq 6\operatorname{tail}(1.6b)$.

Combining the parts. On the other hand, from the anticoncentration bound above, for any a we have

$$\Pr[||q|-a| \geq b] \geq \frac{1}{2}(1-\zeta^{1/4})\left(\operatorname{tail}\left(\frac{b}{1-4\sqrt{\zeta}}\right) - 2\zeta^{1/8}\right).$$

For sufficiently small $\zeta \ll \operatorname{tail}(b)^8$, this is at least $\frac{1}{4}\operatorname{tail}(1.1b)$. So we have

$$\frac{1}{4}\operatorname{tail}(1.1b) \leq 6\operatorname{tail}(1.6b),$$

a contradiction for large enough b. □

Acknowledgements

I would like to thank Laura Mančinska for helpful discussions on the topic of this paper, Tony Short for the "anti-Holevo" terminology, Sandu Popescu for pointing out ref. [17], and Scott Aaronson for notifying me of ref. [3]. This work was supported by EPSRC Early Career Fellowship EP/L021005/1.

A Classical upper bound for the vector in subspace problem

In this appendix we prove a general upper bound on the amount of classical communication required to solve a bounded-error (and therefore at least as hard) variant of the "vector in subspace" problem discussed in Section 1. In this version of the problem, Alice gets an n-qubit quantum state $|\psi\rangle$ and Bob gets a 2-outcome projective measurement $\{M, I-M\}$. Alice and Bob are promised that either $\langle\psi|M|\psi\rangle \geq 2/3$ or $\langle\psi|M|\psi\rangle \leq 1/3$; their task is to output 1 in the first case, and 0 in the second.

We will show that there is a classical randomised protocol for this problem that communicates $O(\sqrt{N})$ bits, where as before we set $N = 2^n$. This implies that any one-way quantum protocol for computing a partial boolean function on n bits can be simulated by a classical protocol communicating $O(\sqrt{N})$ bits [25]. This protocol was proposed by Raz [33] (and subsequently also described by Klartag and Regev [23]), who stated this complexity bound, but no proof has appeared. We first note that we can assume that $\operatorname{tr} M = 2^{n-1}$ without loss of generality; if not, we embed $|\psi\rangle$ and M appropriately in a space of dimension $N' \leq 2N$ such that $\operatorname{tr} M = N'/2$, which does not substantially affect the complexity bounds.

We will need the following technical lemma:

Lemma 7 (Corollary of Bennett et al. [7]). *Let $|\psi\rangle$ be picked from $\mathbb{C}^N$ according to Haar measure on the unit sphere, and let P be the projector onto an r-dimensional subspace of $\mathbb{C}^N$. Then, for any $\delta \geq 0$,*

$$\Pr\left[\langle\psi|P|\psi\rangle \geq (1+\delta)\frac{r}{N}\right] \leq \begin{cases} \exp(-r\delta^2/3) & [0 \leq \delta \leq 1] \\ \exp(-r\delta/3) & [\delta \geq 1] \end{cases}$$

The protocol proceeds as follows. Alice and Bob use shared randomness to specify K quantum states $|\phi_1\rangle, \dots, |\phi_K\rangle$, each picked independently according to Haar measure, for some K (it will turn out that $K = 2^{\Theta(\sqrt{N})}$ suffices). Then Alice finds the state $|\phi_i\rangle$ such that $|\langle\phi_i|\psi\rangle|$ is maximised, and sends the identity of this state to Bob using $\lceil \log_2 K \rceil$ classical bits of communication. Bob then outputs 0 if $\langle\phi_i|M|\phi_i\rangle \leq 1/2$, and 1 otherwise.

We first observe that $|\phi_i\rangle$ can be expressed as

$$|\phi_i\rangle = \epsilon|\psi\rangle + \sqrt{1-\epsilon^2}|\psi^\perp\rangle,$$

where ϵ is a random variable such that $\epsilon = \langle\psi|\phi_i\rangle$, which can be taken to be real and positive such that $\epsilon = |\langle\psi|\phi_i\rangle|$, and $|\psi^\perp\rangle$ is a unit vector distributed uniformly at random in the subspace of states orthogonal to $|\psi\rangle$. This holds because the Haar measure is unitarily invariant, so for each j, $|\phi_j\rangle$ can be picked by choosing its inner product with each vector in an arbitrary orthonormal basis according to a complex Gaussian distribution, then normalising the resulting vector. Once the inner product with $|\psi\rangle$ is fixed, the remaining inner products are independent.

Assume that $\langle\psi|M|\psi\rangle \leq 1/3$ (the case $\langle\psi|M|\psi\rangle \geq 2/3$ is similar, swapping the roles of M and $I-M$). Then the probability that Bob fails to output the correct answer is

$$\Pr_{|\psi^\perp\rangle}\left[\langle\phi_i|M|\phi_i\rangle \geq \frac{1}{2}\right]. \tag{5}$$

We can expand

$$\begin{aligned}\langle\phi_i|M|\phi_i\rangle &= \left(\epsilon\langle\psi| + \sqrt{1-\epsilon^2}\langle\psi^\perp|\right) M \left(\epsilon|\psi\rangle + \sqrt{1-\epsilon^2}|\psi^\perp\rangle\right) \\ &= \epsilon^2\langle\psi|M|\psi\rangle + 2\epsilon\sqrt{1-\epsilon^2}\,\mathrm{Re}(\langle\psi^\perp|M|\psi\rangle) + (1-\epsilon^2)\langle\psi^\perp|M|\psi^\perp\rangle,\end{aligned}$$

and using $\langle\psi|M|\psi\rangle \le 1/3$ and a union bound, can upper-bound (5) as

$$\Pr\left[\langle\phi_i|M|\phi_i\rangle \ge \frac{1}{2}\right] \le \Pr\left[2\epsilon\sqrt{1-\epsilon^2}\,\mathrm{Re}(\langle\psi^\perp|M|\psi\rangle) \ge \frac{\epsilon^2}{12}\right] + \Pr\left[(1-\epsilon^2)\langle\psi^\perp|M|\psi^\perp\rangle \ge \frac{1}{2} - \frac{5\epsilon^2}{12}\right].$$

We bound each of the remaining two terms separately. For the first term, we use

$$\begin{aligned}\Pr\left[2\epsilon\sqrt{1-\epsilon^2}\,\mathrm{Re}(\langle\psi^\perp|M|\psi\rangle) \ge \frac{\epsilon^2}{12}\right] &\le \Pr\left[|\langle\psi^\perp|M|\psi\rangle| \ge \frac{\epsilon}{24}\right] \\ &= \Pr\left[\frac{|\langle\psi^\perp|(I-|\psi\rangle\langle\psi|)M|\psi\rangle|^2}{\|(I-|\psi\rangle\langle\psi|)M|\psi\rangle\|^2} \ge \left(\frac{\epsilon}{24\|(I-|\psi\rangle\langle\psi|)M|\psi\rangle\|}\right)^2\right] \\ &\le \exp(-((\epsilon/24)^2(N-1)-1)/3)\end{aligned}$$

using Lemma 7 with $r=1$, $\delta = (\epsilon/24)^2(N-1)-1$, and assuming that $\epsilon \gg N^{-1/2}$. For the second term, we use

$$\langle\psi^\perp|M|\psi^\perp\rangle \le \langle\psi^\perp|\,\mathrm{supp}((I-|\psi\rangle\langle\psi|)M(I-|\psi\rangle\langle\psi|))|\psi^\perp\rangle,$$

where $\mathrm{supp}(X)$ denotes the projector onto the support of X, and

$$N/2 = \mathrm{rank}(M) \ge \mathrm{rank}((I-|\psi\rangle\langle\psi|)M(I-|\psi\rangle\langle\psi|)) = \mathrm{rank}((I-|\psi\rangle\langle\psi|)M) \ge \mathrm{rank}(M)-1 = N/2-1,$$

together with Lemma 7 and the bound

$$\frac{1}{1-\epsilon^2}\left(\frac{1}{2}-\frac{5\epsilon^2}{12}\right) = \frac{6-5\epsilon^2}{12(1-\epsilon^2)} \ge \frac{1}{2}+\frac{\epsilon^2}{12}$$

to obtain

$$\Pr\left[\langle\psi^\perp|M|\psi^\perp\rangle \ge \frac{1}{1-\epsilon^2}\left(\frac{1}{2}-\frac{5\epsilon^2}{12}\right)\right] \le \exp(-CN\epsilon^4)$$

for some universal constant C. It is therefore sufficient to have $\epsilon = \Omega(N^{-1/4})$ in order for the sum of the two probabilities to be bounded by an arbitrarily small constant. All that remains is to determine how large K needs to be to achieve this. Recall that ϵ was the largest absolute inner product between any of the random states $|\phi_j\rangle$ and the fixed state $|\psi\rangle$. For any $0 \le x \le 1$, we have

$$\Pr_{|\phi_j\rangle}[|\langle\phi_j|\psi\rangle|^2 \ge x] = (1-x)^{N-1}.$$

To see this, first note that the distribution obtained by measuring a Haar-random N-dimensional state $|\phi_i\rangle$ in an arbitrary orthonormal basis is uniform in the probability simplex [35]. Geometrically, this corresponds to a standard simplex in $N-1$ spatial dimensions. Truncating this simplex by restricting one coordinate from the range $[0,1]$ to the range $[x,1]$ gives a geometrically similar simplex, whose volume must therefore be the volume of the original simplex multiplied by $(1-x)^{N-1}$. So, for any $0 < D < N^{1/2}$,

$$\Pr_{|\phi_j\rangle}[|\langle\phi_j|\psi\rangle|^2 \ge DN^{-1/2}] = (1-DN^{-1/2})^{N-1} \ge (e^{-2DN^{-1/2}})^{N-1} \ge e^{-2D\sqrt{N}}.$$

It is therefore sufficient to choose $K = 2^{O(\sqrt{N})}$ for there to exist some $i \in \{1,\dots,K\}$ such that $|\langle\phi_i|\psi\rangle| \ge N^{-1/4}$ with high probability, corresponding to $O(\sqrt{N})$ bits of communication being required to specify $|\phi_i\rangle$.

B Direct lower bound on the classical query complexity of Fourier sampling

In this appendix we give a simpler direct proof of Corollary 2:

Corollary 2 (restated). *For sufficiently small constant $\epsilon > 0$, Fourier Sampling on N input bits requires $\Omega(N)$ classical queries.*

Recall that in this problem we are given query access to a function $h : \{0,1\}^n \to \{\pm 1\}$ (equivalently, to an arbitrary string of $N = 2^n$ ± 1's), and are asked to output a sample from any distribution $\widetilde{p}_h$ such that $\|\widetilde{p}_h - p_h\|_1 \le \epsilon$, where

$$p_h(s) = \left(\frac{1}{2^n} \sum_{x \in \{0,1\}^n} (-1)^{s \cdot x} h(x) \right)^2 .$$

Proof. Following the same argument as at the start of Section 2, from any classical algorithm which solves Fourier Sampling with k queries, we can obtain an algorithm of the following form. Given oracle access to a bit-string $x \in \{\pm 1\}^N$, the algorithm makes k queries to elements of x and accepts with probability q_x such that

$$\mathbb{E}_x \left[\left| q_x - \left(\frac{1}{N} \sum_i x_i \right)^2 \right| \right] \le \frac{\epsilon}{N},$$

where x is uniformly random. Because of this randomness, we can assume that the algorithm deterministically queries the first k bits of x and its acceptance probability depends only on these. Thus, splitting x into a length-k "queried" string y and a length-m "unqueried" string z such that $k + m = N$, we want to find a sequence of numbers q_y to minimise

$$\mathbb{E}_{y,z} \left[\left| q_y - \left(\frac{1}{N} \left(\sum_i y_i + \sum_j z_j \right) \right)^2 \right| \right].$$

For convenience, take out a factor of N^2, rescaling q_y appropriately. Then we want to minimise

$$\frac{1}{N^2} \mathbb{E}_{y,z} \left[\left| q_y - \left(\sum_i y_i + \sum_j z_j \right)^2 \right| \right] = \frac{1}{N^2} \mathbb{E}_{y,z} \left[\left| q_y - \left(\sum_i y_i \right)^2 - \left(\sum_j z_j \right)^2 - 2 \left(\sum_i y_i \right) \left(\sum_j z_j \right) \right| \right].$$

Shifting q_y by $(\sum_i y_i)^2$ (which we are free to do as q_y is an arbitrary function of y), and using the reverse triangle inequality, we can lower-bound this expression by

$$\frac{1}{N^2} \mathbb{E}_{y,z} \left[\left| q_y - \left(\sum_j z_j \right)^2 \right| \right] - \frac{2}{N^2} \mathbb{E}_{y,z} \left[\left| \left(\sum_i y_i \right) \left(\sum_j z_j \right) \right| \right].$$

The random variables y and z are independent and uniformly distributed; and we have $\mathbb{E}_y [|\sum_i y_i|] = O(\sqrt{k})$, $\mathbb{E}_z [|\sum_i z_i|] = O(\sqrt{m})$. So the second term is at most $O(\sqrt{km}/N^2)$ in magnitude; and the first term now does not depend on y. It remains to bound this term, i.e. to lower-bound

$$\min_q \frac{1}{N^2} \mathbb{E}_z \left[\left| q - \left(\sum_j z_j \right)^2 \right| \right].$$

We split into cases. If $q \leq m$, this expression is lower-bounded by

$$\frac{1}{N^2} m \Pr_z \left[\left(\sum_j z_j \right)^2 \geq 2m \right] \geq \frac{Cm}{N^2}$$

for some constant C; similarly, if $q \geq m$, we obtain a lower bound of

$$\frac{1}{N^2} \frac{m}{2} \Pr_z \left[\left(\sum_j z_j \right)^2 \leq \frac{m}{2} \right] \geq \frac{C'm}{N^2}$$

for some constant C'. Therefore, the whole expression is lower-bounded by $(Cm - D\sqrt{km})/N^2 = (m/N^2)(C - D\sqrt{N/m - 1})$ for some universal constants C and D. For small enough $\epsilon > 0$, there exists a universal constant $D' < 1$ such that, for all $m \geq D'N$, this is strictly greater than ϵ/N. □

References

[1] S. Aaronson. The learnability of quantum states. *Proceedings of the Royal Society A*, 463:2088, 2007. quant-ph/0608142.

[2] S. Aaronson and A. Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. In *Proc. 47th Annual ACM Symp. Theory of Computing*, pages 307–316, 2015. arXiv:1411.5729.

[3] S. Aaronson and L. Chen. Complexity-theoretic foundations of quantum supremacy experiments, 2016. arXiv:1612.05903.

[4] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. Dense quantum coding and quantum finite automata. *J. ACM*, 49(4):496–511, 2002. quant-ph/9804043.

[5] Z. Bar-Yossef, T. S. Jayram, and I. Kerenidis. Exponential separation of quantum and classical one-way communication complexity. *SIAM J. Comput.*, 38(1):366–384, 2008.

[6] P. Beame, T. Pitassi, N. Segerlind, and A. Wigderson. A strong direct product theorem for corruption and the multiparty communication complexity of disjointness. *Computational Complexity*, 15:391–432, 2007.

[7] C. H. Bennett, P. Hayden, D. Leung, P. Shor, and A. Winter. Remote preparation of quantum states. *IEEE Trans. Inform. Theory*, 51(1):56–74, 2005. quant-ph/0307100.

[8] E. Bernstein and U. Vazirani. Quantum complexity theory. *SIAM J. Comput.*, 26(5):1411–1473, 1997.

[9] G. Brassard, R. Cleve, and A. Tapp. Cost of exactly simulating quantum entanglement with classical communication. *Phys. Rev. Lett.*, 83(9):1874–1877, 1999. quant-ph/9901035.

[10] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner. Bell nonlocality. *Rev. Mod. Phys.*, 86:419, 2014. arXiv:1303.2849.

[11] H. Buhrman, R. Cleve, S. Massar, and R. de Wolf. Non-locality and communication complexity. *Rev. Mod. Phys.*, 82(1):665–698, 2010. arXiv:0907.3584.

[12] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *Proc. 30th Annual ACM Symp. Theory of Computing*. ACM Press, 1998. quant-ph/9802040.

[13] N. Cerf, N. Gisin, and S. Massar. Classical teleportation of a quantum bit. *Phys. Rev. Lett.*, 84:2521, 1999. quant-ph/9906105.

[14] A. Chakrabarti, R. Kondapally, and Z. Wang. Information complexity versus corruption and applications to orthogonality and Gap-Hamming. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2012)*, pages 483–494, 2012. arXiv:1205.0968.

[15] A. Chakrabarti and O. Regev. An optimal lower bound on the communication complexity of Gap-Hamming-Distance. *SIAM J. Comput.*, 41(5):1299–1317, 2012. arXiv:1009.3460.

[16] D. Deutsch and R. Jozsa. Rapid solution of problems by quantum computation. *Proc. Roy. Soc. London Ser. A*, 439(1907):553–558, 1992.

[17] E. Galvão and L. Hardy. Substituting a qubit for an arbitrarily large number of classical bits. *Phys. Rev. Lett.*, 90:087902, 2003. quant-ph/0110166.

[18] D. Gavinsky. Classical interaction cannot replace a quantum message. In *Proc. 40th Annual ACM Symp. Theory of Computing*, pages 95–102, 2008. quant-ph/0703215.

[19] D. Gavinsky, J. Kempe, I. Kerenidis, R. Raz, and R. de Wolf. Exponential separations for one-way quantum communication complexity, with applications to cryptography. *SIAM J. Comput.*, 38(5):1695–1708, 2008. quant-ph/0611209.

[20] M. Göös and T. Watson. Communication complexity of set-disjointness for all probabilities. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2014)*, pages 721–736, 2014.

[21] P. Hayden, D. Leung, P. Shor, and A. Winter. Randomizing quantum states: Constructions and applications. *Comm. Math. Phys.*, 250(2):371–391, 2004. quant-ph/0307104.

[22] A. S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problemy Peredachi Informatsii*, 9(3):3–11, 1973. English translation *Problems of Information Transmission*, vol. 9, pp. 177-183, 1973.

[23] B. Klartag and O. Regev. Quantum one-way communication can be exponentially stronger than classical communication. In *Proc. 43rd Annual ACM Symp. Theory of Computing*, pages 31–40, 2011. arXiv:1009.3640.

[24] H. Klauck. Rectangle size bounds and threshold covers in communication complexity. In *Proc. 18th Annual IEEE Conf. Computational Complexity*, pages 118–134, 2003. cs/0208006.

[25] I. Kremer. Quantum communication. Master's thesis, Hebrew University, 1995.

[26] I. Kremer, N. Nisan, and D. Ron. On randomized one-round communication complexity. *Computational Complexity*, 8:21–49, 1999.

[27] E. Kushilevitz and N. Nisan. *Communication Complexity.* Cambridge University Press, 1997.

[28] S. Laplante, M. Laurière, Alexandre Nolin, Jérémie Roland, and Gabriel Senno. Robust bell inequalities from communication complexity, 2016. `arXiv:1606.09514`.

[29] A. Montina. Exponential communication gap between weak and strong classical simulations of quantum communication. *Phys. Rev. A*, 87:042331, 2013. `arXiv:1301.3452`.

[30] A. Montina, M. Pfaffhauser, and S. Wolf. Communication complexity of channels in general probabilistic theories. *Phys. Rev. Lett.*, 111:160502, 2013. `arXiv:1301.4441`.

[31] A. Montina and S. Wolf. Lower bounds on the communication complexity of two-party (quantum) processes. In *Proc. 2014 IEEE International Symposium on Information Theory*, pages 1484–1488, 2014. `arXiv:1401.4126`.

[32] A. Nayak and F. Wu. The quantum query complexity of approximating the median and related statistics. In *Proc. 31st Annual ACM Symp. Theory of Computing*, pages 384–393, 1999. `quant-ph/9804066`.

[33] R. Raz. Exponential separation of quantum and classical communication complexity. In *Proc. 31st Annual ACM Symp. Theory of Computing*, pages 358–367, 1999.

[34] A. Sherstov. The communication complexity of gap Hamming distance. *Theory of Computing*, 8:197–208, 2012.

[35] S. Sýkora. Quantum theory and the Bayesian inference problems. *Journal of Statistical Physics*, 11(1):17–27, 1974.

[36] B. Toner and D. Bacon. Communication cost of simulating Bell correlations. *Phys. Rev. Lett.*, 91:187904, 2003. `quant-ph/0304076`.

[37] T. Vidick. A concentration inequality for the overlap of a vector on a large set, with application to the communication complexity of the gap-Hamming-distance problem. *Chicago Journal of Theoretical Computer Science*, 2012(1):1–12, 2012.

[38] A. Yao. Lower bounds by probabilistic arguments. In *Proc. 24th Annual Symp. Foundations of Computer Science*, pages 420–428, 1983.