Quantum states cannot be transmitted efficiently classically

Ashley Montanaro[*]

October 24, 2017

Abstract

We show that any classical two-way communication protocol with shared randomness 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), up to constant accuracy, must transmit at least $\Omega(2^n)$ bits. This lower bound is optimal and matches the complexity of a simple protocol based on discretisation using an ϵ-net. The proof 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 easy corollaries. First, a sampling problem which can be solved with one quantum query to the input, but which requires $\Omega(N)$ 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 $\Omega(2^n)$ bits.

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 [23], 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 [35] 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

---

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

on measurement outcomes would be almost the same. Therefore, the most we can reasonably ask 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 $\|p_M(|\psi\rangle) - \widetilde{p}_M(|\psi\rangle)\|_1 \le \epsilon$, for some $\epsilon > 0$, where $\|\cdot\|_1$ is the ℓ_1 distance $\|v\|_1 = \sum_i |v_i|$. Call this task *Distributed Quantum Sampling*. Distributed Quantum 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.

Distributed Quantum 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 [22], i.e. a set of states $\{|\psi_i\rangle\}$ such that, for all $|\psi\rangle$, $\||\psi\rangle\langle\psi| - |\psi_i\rangle\langle\psi_i|\|_1 \le \epsilon$ 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 [22], 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 [22] 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 [12]. One of the first results in this area was work of Buhrman, Cleve and Wigderson [13] which implies that, in our terminology, Distributed Quantum 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 [36] 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. [20] 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 Distributed Quantum Sampling requires $\Omega(\sqrt[3]{2^n})$ classical bits of communication between Alice and Bob [25], even if the communication is allowed to be two-way. This improved a previous result of Gavinsky [19], which implied that Distributed Quantum Sampling requires $\Omega(\sqrt[8]{2^n}/\sqrt{n})$ bits of classical two-way communication. The lower bound of [25] 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

---

[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 $O(2^n \log(2^n/\epsilon))$.

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 exact quantum protocols where Bob has two possible outputs [27].

Thus, all these results (and more discussed in Section 1.5 below) leave open a natural question: could there exist a non-trivial classical protocol for Distributed Quantum 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 [36] 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 Distributed Quantum Sampling problem? In the author's opinion, the interest in this question goes beyond simply finding a tight bound for this problem that would improve the best known $\Omega(\sqrt[3]{2^n})$ lower bound. On a fundamental, conceptual level, the question asks: 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 Distributed Quantum 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 and the parties are allowed shared randomness. 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 [23], but $\Omega(2^n)$ classical bits are required to store n qubits.

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

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

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

- 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 $\|\widetilde{p}_{fg} - p_{fg}\|_1 \le \epsilon$, 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:

---

[2]This result was stated in [36] but the proof has not appeared. We include a proof in Appendix A.

Theorem 1. *There exist universal constants $\epsilon, \gamma > 0$ such that, for sufficiently large N, any two-way classical communication protocol for Distributed Fourier Sampling with shared randomness and inaccuracy ϵ must communicate at least γN bits.*

Theorem 1 implies an optimal lower bound on Distributed Quantum Sampling for $\epsilon = \Omega(1)$, matching the general upper bound of $O(N \log(1/\epsilon))$ bits. It remains open to prove a tight bound when $\epsilon = o(1)$. However, note that a lower bound of $\Omega(N + \log(1/\epsilon))$ does hold, because an $\Omega(\log 1/\epsilon)$ bound is straightforward. If we let Alice's state be $\alpha|0\rangle + \beta|1\rangle$, and Bob's measurement is fixed to be a computational basis measurement, sampling from the corresponding probability distribution up to inaccuracy ϵ is effectively the same as transmitting $|\alpha|^2$ up to inaccuracy $O(\epsilon)$. This requires transmitting $\Omega(\log 1/\epsilon)$ 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 query algorithm 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 [29]. 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[3]. 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

---

[3]Note that, unlike in the setting of quantum query algorithms, it is not necessary for Alice and Bob to send each other their query indices. The classical algorithm can be expressed as a distribution over deterministic decision trees. Alice and Bob choose which deterministic tree to implement using shared randomness, and from that point onwards, the values of the queried bits completely determine the algorithm's behaviour.

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.

Following the completion of an initial version of this work, I learned that Scott Aaronson and Lijie Chen have recently obtained an independent proof of Corollary 2 [3]. (Their proof technique is compared with the present one in Section 1.3 below.) Also note that it is much simpler to prove Corollary 2 for deterministic classical algorithms, i.e. ones that choose which queries to make via a deterministic decision tree, then sample from some distribution depending on the results of the queries. A proof for this special case is included in Appendix B.

Nonlocality problems. Using a standard mapping between communication protocols and entanglement [12], 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 [39].

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 communication protocol for DDFS approximating the output distribution up to ℓ_1 inaccuracy ϵ with cost c gives a protocol for Distributed Fourier Sampling up to inaccuracy ϵ with cost $c + n$. 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 $\tilde{q}_{fg}(s, t)$, 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

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

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

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 inaccuracy using $O(N)$ bits of communication, via a similar ϵ-net construction as for Distributed Quantum 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 [17], 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 [13, 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. The starting point is to observe that, 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

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

for most inputs f, g. More specifically, there is 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 uniformly random f and g is at most ϵ/N. A similar idea was used in [2, 3] in the setting of query complexity.

Next, we show that 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. It is convenient to now switch notation and consider the problem of accepting with probability close to $(\langle x,y \rangle/N)^2$ for strings $x,y \in \{\pm 1\}^N$. The first challenge in proving a lower bound is that it is hard to reason about protocols for sampling problems. In particular, one cannot necessarily assume that the existence of a bounded-error randomised protocol for a sampling problem implies the existence

6

of a deterministic protocol for the same problem that is accurate on most inputs[4]. This is by contrast with the case of communication complexity of functional (or relational) problems [29], where this connection is standard and is a typical first step in the proof of lower bounds.

We instead need to reason directly about the acceptance probability of any low-communication protocol P. Our high-level strategy is to find three distributions $\mathcal{D}_-$, $\mathcal{D}_0$, $\mathcal{D}_+$ on inputs such that we can lower-bound the acceptance probability p_0 of P under $\mathcal{D}_0$ in terms of a weighted sum of its acceptance probabilities p_-, p_+ under $\mathcal{D}_-$, $\mathcal{D}_+$. Assuming that P is accurate allows us to find an upper bound on each of these two acceptance probabilities (and hence an upper bound on p_0), as well as a lower bound on p_0. If the upper bound is lower than the lower bound, we have obtained a contradiction and there hence cannot exist an accurate low-communication protocol.

To implement this strategy mathematically, we use the fact that a communication protocol can be understood in terms of rectangles, i.e. subsets of Alice and Bob's inputs of the form $R = A \times B$. First, it is not difficult to see that the only rectangles that contribute appreciably to the protocol's average acceptance probability under some distribution are those that are large under that distribution. Next we want to show that, for any large rectangle that contributes substantially to the acceptance probability under some distribution $\mathcal{D}_0$, that rectangle must contribute even more substantially to the acceptance probabilities under some other distributions $\mathcal{D}_-$, $\mathcal{D}_+$.

At this point, we can make a connection to a remarkably powerful previous technique used to determine the communication complexity of the so-called gap-Hamming and gap-orthogonality problems [16, 15, 40, 37]. In the gap-Hamming problem Alice and Bob are each given strings $x, y \in \{\pm 1\}^N$, under the promise that either $\langle x, y \rangle \leq -\sqrt{N}$ or $\langle x, y \rangle \geq \sqrt{N}$, where as above $\langle x, y \rangle = \sum_i x_i y_i$; their goal is to determine which is the case. It was first shown by Chakrabarti and Regev that the communication complexity of the gap-Hamming problem is $\Omega(N)$ [16]. The key technical step of their bound involved giving three distributions ξ_0, ξ_-, ξ_+ on pairs of strings $x, y \in \{\pm 1\}^N$ such that, for any large rectangle R,

$$\frac{1}{2}\left(\xi_-(R) + \xi_+(R)\right) \geq \frac{2}{3}\xi_0(R). \tag{3}$$

The distributions ξ_- and ξ_+ are concentrated on strings with relatively large and small (respectively) Hamming distance; indeed, under these distributions x is uniformly random, while for each i, $y_i = x_i$ with probability $(1 \pm p)/2$, for some small p. ξ_0 is just the uniform distribution. Thus (3) can be seen as an anticoncentration bound: any large rectangle (i.e. any rectangle that contains many pairs (x, y) with respect to the uniform distribution) must contain many pairs (x, y) such that $|\langle x, y \rangle|$ is large.

This appears to give us exactly what we need to complete the proof. However, there are some difficulties left to surmount. First, the corresponding acceptance probabilities of a good sampler under the distributions ξ_0, ξ_-, ξ_+ used in [16] are not the ones that we need to use to prove the desired contradiction. To address this, Alice and Bob randomly flip some of their input bits in a suitably correlated way, which corresponds to shifting the expected acceptance probability $(\langle x, y \rangle/N)^2$ under each of the distributions. Even after doing this, the bound (3) turns out to not be enough to imply a contradiction. We instead need a skewed version of the bound:

$$\frac{1}{2}\left(e^s \xi_-(R) + e^{-s} \xi_+(R)\right) \geq \frac{2}{3}\xi_0(R) \tag{4}$$

for large rectangles R and arbitrary $s \in \mathbb{R}$. Intuitively, in this expression we are improving the lower bound in terms of ξ_+, at the expense of making the bound in terms of ξ_- worse. It turns

---

[4]A previous version of this paper claimed incorrectly that such an implication did hold.

out to indeed be possible to prove such a generalised bound based on the techniques of [16][5]. This generalised bound could find applications elsewhere.

The final issue to deal with is apparently a technicality, but perhaps an important technicality. We have from the accuracy constraint on the Distributed Fourier Sampling protocol that the expected difference between the protocol's acceptance probability and $(\langle x, y \rangle / N)^2$ over uniformly random x and y is at most ϵ/N. However, for the above bounds to be meaningful we need the protocol to also have a comparable level of accuracy when (x, y) are picked from the distributions ξ_+, ξ_-. But the distributions we use are relatively far from the uniform distribution in total variation distance, implying that an expected error of ϵ/N under this distribution may not translate into a small error under these distributions. To address this, Alice and Bob map their inputs (x, y) to new inputs (x', y') that are larger by a constant factor but preserve the inner product $\langle x, y \rangle$. This allows us to show that all probabilities in the corresponding distributions $\xi'_\pm$ on the larger inputs are at most a constant multiple (depending on ϵ) of their corresponding probabilities under the uniform distribution, so the accuracy bound can be translated across.

We finally remark on connections between the proof techniques of the tight bound given here on the communication complexity of Distributed Fourier Sampling, and the tight bound of Aaronson and Chen on the query complexity of Fourier Sampling [3]. At a high level, Aaronson and Chen's strategy is also to prove hardness of accepting with probability close to $(\langle x, y \rangle / N)^2$, based on showing contradictory bounds on the acceptance probability of any classical algorithm that makes few queries, under three different input distributions. As query algorithms are much more restrictive than communication protocols, this allows the acceptance probability of an algorithm in this setting to be written as a weighted sum of certain functions of binomial coefficients. To complete the proof strategy of [3] it is then sufficient to carry out some technical calculations to bound these coefficients. In the communication complexity context, it seems necessary to prove a more general bound of the form of (4), which comes with significant technical complications. For example, it does not seem obvious how to prove a variant of (4) corresponding to the distributions used in [3].

1.4 Barriers to use of previous results

It is instructive to check why we cannot just use existing communication complexity results to easily prove a lower bound on Distributed Fourier Sampling. First, by the $O(\sqrt{2^n})$ classical upper bound of Raz [36] 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 gap-orthogonality problem [37, 15], for which an $\Omega(N)$ lower bound is known in the case where acceptance and rejection probabilities are separated by an *additive* constant. So 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 [21] showed that communication lower bounds of this form can be proven using the corruption method [41, 6, 26], an important lower bound technique in

---

[5]For experts, we remark that the ultimate reason for this is that the bound of [16] hinges on the behaviour of the function $\cosh(z)$, while here we need to understand the function $\cosh(z - s)$, which is very similar; and that it might also be possible to use a suitably generalised and tightened version of bounds proven in [15] to obtain a similar result.

communication complexity, used in particular by Sherstov to prove his $\Omega(N)$ lower bound for gap-orthogonality [37]. 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 [21]. 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.

1.5   Other related prior work

The Distributed Quantum Sampling problem has also been studied in the physics literature, where it is sometimes termed "classical teleportation". Toner and Bacon [39] showed that, in the case where Bob makes a projective measurement, Distributed Quantum 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 [14]. Montina [32] 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 [28]). Montina, Pfaffhauser and Wolf [33], and Montina and Wolf [34], 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 [11, 12] for surveys.

Galvão and Hardy [18] considered a variant of the Distributed Quantum 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. (See [10] for a related result applied to the study of non-classical temporal correlations.) 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.

We finally remark that, using a connection between the corruption bound and Bell inequalities, the gap-orthogonality problem was recently used by Laplante et al. [30] to give inefficiency-resistant Bell inequalities with large violations.

2 Preliminary definitions

We use standard definitions from communication complexity [29]. We consider the setting of two-way communication protocols with shared randomness, with two parties (Alice and Bob), where Alice's input is often denoted $x \in X$, and Bob's input is denoted $y \in Y$. A *rectangle* $R = A \times B$ is a product subset of Alice and Bob's inputs ($A \subseteq X$, $B \subseteq Y$). We usually consider sampling problems, where Alice and Bob are given the pair (x, y) and asked to output a sample from some predetermined distribution $\mathcal{D}_{xy}$, up to some predetermined accuracy.

We sometimes use the convention $[Q]$ for a quantity that evaluates to 1 if Q is true, and 0 otherwise, while $[a, b]$ denotes the interval $a \leq z \leq b$. We will use the following family of distributions on correlated pairs (x, y) throughout:

Definition 1 (cf. [16]). *For $p \in [-1, 1]$, let ξ_p^N denote the distribution on $(x, y) \in \{\pm 1\}^N \times \{\pm 1\}^N$ obtained by picking $x \in \{\pm 1\}^N$ uniformly at random, then picking y by setting $y_i = x_i$ with probability $(1+p)/2$, and $y_i = -x_i$ with probability $(1-p)/2$. Note that ξ_0^N is the uniform distribution on $\{\pm 1\}^N \times \{\pm 1\}^N$. When N is implicit, we often write $\xi_p := \xi_p^N$.*

3 Proof of main result

We now describe the proof of Theorem 1 in detail. We first state the technical lemmas from which the main theorem follows, and give its proof assuming these lemmas. The lemmas themselves are proved afterwards.

The first lemma is a characterisation of shared-randomness classical communication protocols in terms of rectangles. Note that similar bounds are common in the communication complexity literature (see e.g. [24]), but are usually stated only for decision problems. As we are interested in the more general setting of sampling problems, we include a proof in Section 4.

Lemma 4 (Characterisation of communication protocols). *Let P be a communication protocol with two outcomes (accept and reject) and shared randomness, such that P communicates at most c bits. Then there exists a non-negative function $\rho(R)$ of rectangles R such, for any distribution μ on pairs $(x, y) \in X \times Y$,*

$$\Pr[P \text{ accepts } (x, y)] = \sum_R \rho(R)\mu(R) = \eta + \sum_{R, \mu(R) \geq 2^{-2c}} \rho(R)\mu(R)$$

for some $\eta \in [0, 2^{-c})$, where the probability is taken over both $(x, y) \sim \mu$ and the internal randomness of P.

Lemma 5 (From Fourier sampling to two-outcome protocols). *Assume that, for all sufficiently large N, there exists a protocol for Distributed Fourier Sampling with inaccuracy ϵ that communicates at most γN bits. Fix $b = O(1)$. Then there exists a protocol that communicates at most $2\gamma N$ bits and accepts with probability $\widetilde{p}_{xy}$ on input $(x, y) \in \{\pm 1\}^N \times \{\pm 1\}^N$ such that:*

$$\mathbb{E}_{(x,y) \sim \xi_{-\sqrt{b/N}}} [\widetilde{p}_{xy}] \in \left[\frac{1 - \epsilon'}{4N}, \frac{1 + \epsilon'}{4N}\right], \quad \mathbb{E}_{(x,y) \sim \xi_0} [\widetilde{p}_{xy}] \in \left[\frac{(b+1)\beta - \epsilon'}{4N}, \frac{(b+1)\beta + \epsilon'}{4N}\right],$$

$$\mathbb{E}_{(x,y)\sim\xi_{\sqrt{b/N}}}[\widetilde{p}_{xy}] \in \left[\frac{(4b+1)\beta - \epsilon'}{4N}, \frac{(4b+1)\beta + \epsilon'}{4N}\right],$$

where $\epsilon' = e^{Cb}\epsilon$ for some universal constant C, and $\beta \in [1 - o(1), 1]$.

Lemma 6 (Skewed anticoncentration bound for large rectangles)**.** *For all $b > 0$ there exists $\delta > 0$ such that for large enough N, all rectangles $R \subseteq \{\pm 1\}^N \times \{\pm 1\}^N$ such that $\xi_0(R) \geq 2^{-\delta N}$, and all s,*

$$\frac{1}{2}\left(e^s \xi_{-\sqrt{b/N}}(R) + e^{-s}\xi_{\sqrt{b/N}}(R)\right) \geq \frac{2}{3}\xi_0(R).$$

Theorem 1 (restated). *There exist universal constants $\epsilon, \gamma > 0$ such that, for sufficiently large N, any two-way classical communication protocol for Distributed Fourier Sampling with shared randomness and inaccuracy ϵ must communicate at least γN bits.*

Proof. Assume towards a contradiction that for all sufficiently large N there exists a protocol for Distributed Fourier Sampling with inaccuracy ϵ that communicates at most γN bits, for arbitrarily small constants $\epsilon, \gamma > 0$. Let P be the corresponding two-outcome communication protocol that communicates $c \leq 2\gamma N$ bits and has acceptance probability bounds given by Lemma 5, for some choice of $b = O(1)$ to be determined. By Lemma 4, P in turn implies the existence of a non-negative function $\rho(R)$ of rectangles R such that

$$p_0 := \Pr_{(x,y)\sim\xi_0}[P \text{ accepts } (x,y)] = \eta + \sum_{R, \xi_0(R) \geq 2^{-2c}} \rho(R)\xi_0(R) \tag{5}$$

for some $\eta \in [0, 2^{-c})$. By Lemma 5,

$$p_0 \geq \frac{(b+1)\beta - \epsilon'}{4N},$$

where $\epsilon' = e^{Cb}\epsilon$ and $\beta = 1 - o(1)$. By (5) and Lemma 6, assuming that $\gamma \leq \delta/4$, for any $s \in \mathbb{R}$

$$
\begin{aligned}
p_0 &\leq 2^{-c} + \sum_{R, \xi_0(R) \geq 2^{-2c}} \rho(R)\frac{3}{4}\left(e^s\xi_{-\sqrt{b/N}}(R) + e^{-s}\xi_{\sqrt{b/N}}(R)\right) \\
&\leq 2^{-c} + \frac{3}{4}\left(e^s\sum_R \rho(R)\xi_{-\sqrt{b/N}}(R) + e^{-s}\sum_{R'}\rho(R')\xi_{\sqrt{b/N}}(R')\right) \\
&\leq 2^{-c} + \frac{3}{4}\left(e^s \Pr_{(x,y)\sim\xi_{-\sqrt{b/N}}}[P \text{ accepts } (x,y)] + e^{-s}\Pr_{(x,y)\sim\xi_{\sqrt{b/N}}}[P \text{ accepts } (x,y)]\right) \\
&\leq 2^{-c} + \frac{3}{4}\left(e^s\frac{1 + \epsilon'}{4N} + e^{-s}\frac{(4b+1)\beta + \epsilon'}{4N}\right),
\end{aligned}
$$

where the final inequality is Lemma 5. Choosing, for example, $b = 10$, $s = 2$, we get that for some universal constant C,

$$\frac{11}{4N} \leq \frac{3}{16N}\left(e^2 + \frac{41}{e^2}\right) + C\frac{\epsilon'}{N} + o(1/N) \leq \frac{10}{4N} + C\frac{\epsilon'}{N} + o(1/N),$$

a contradiction for small enough $\epsilon = \Omega(1)$ and large enough N. Therefore such a protocol cannot exist. $\qquad\square$

4 Proofs of communication complexity lemmas

We first prove a mathematical characterisation of communication protocols in terms of rectangles.

Lemma 4 (restated). *Let P be a communication protocol with two outcomes (accept and reject) and shared randomness, such that P communicates at most c bits. Then there exists a non-negative function $\rho(R)$ of rectangles R such, for any distribution μ on pairs $(x, y) \in X \times Y$,*

$$\Pr[P \ accepts \ (x, y)] = \sum_R \rho(R)\mu(R) = \eta + \sum_{R, \mu(R) \geq 2^{-2c}} \rho(R)\mu(R)$$

for some $\eta \in [0, 2^{-c})$, where the probability is taken over both $(x, y) \sim \mu$ and the internal randomness of P.

Proof. For any (x, y), we have

$$\Pr[P \text{ accepts } (x, y)] = \sum_D \pi(D)[D \text{ accepts } (x, y)] = \sum_D \pi(D) \sum_{R \in \mathcal{R}_D} [(x, y) \in R],$$

where $\pi(D)$ is a distribution on deterministic protocols D, each of which corresponds to a disjoint set of 1-rectangles $\mathcal{R}_D$. Taking the expectation over μ,

$$
\begin{aligned}
\Pr_{(x,y) \sim \mu, D \sim \pi}[P \text{ accepts } (x, y)] &= \sum_D \pi(D) \sum_{R \in \mathcal{R}_D} \mu(R) \\
&= \sum_D \pi(D) \sum_{R \in \mathcal{R}_D, \mu(R) < 2^{-2c}} \mu(R) + \sum_D \pi(D) \sum_{R \in \mathcal{R}_D, \mu(R) \geq 2^{-2c}} \mu(R) \\
&= \eta + \sum_D \pi(D) \sum_{R \in \mathcal{R}_D, \mu(R) \geq 2^{-2c}} \mu(R),
\end{aligned}
$$

where we infer that $0 \leq \eta < 2^{-c}$ by using $|\mathcal{R}_D| \leq 2^c$ for all D and $\sum_D \pi(D) = 1$. The claim follows by taking $\rho(R) = \sum_{D, R \in \mathcal{R}_D} \pi(D)$. $\qquad\square$

Next we prove a lemma that maps between a distributed Fourier sampler and a two-outcome communication protocol whose acceptance probability obeys certain bounds. In order to prove the lemma, we need the following technical claims, whose proofs are deferred to the very end:

Fact 7.
$$\mathbb{E}_{(x,y) \sim \xi_p}\left[\left(\frac{\langle x, y \rangle}{N}\right)^2\right] = \frac{1}{N} + \left(1 - \frac{1}{N}\right)p^2.$$

Fact 8. *Let ξ_p' be the distribution on pairs $(x', y') \in \{\pm 1\}^{2N}$ formed by choosing $(x, y) \sim \xi_p$, then setting $x' = \sigma(x, x'')$, $y' = \sigma(y, y'')$, where $x'', y'' \in \{\pm 1\}^N$ are fixed strings such that $\langle x'', y'' \rangle = 0$, and $\sigma \in S_{2N}$ is a random permutation of the indices of the strings. Then there exists a universal constant C such that, for sufficiently large N and any Δ, and $p \in [-0.01, 0.01]$,*

$$\Pr_{(x', y') \sim \xi_p'}[\langle x', y' \rangle = \Delta] \leq C e^{2p^2 N} \Pr_{(x', y') \sim \xi_0^{2N}}[\langle x', y' \rangle = \Delta].$$

Lemma 5 (restated). *Assume that, for all sufficiently large N, there exists a protocol for Distributed Fourier Sampling with inaccuracy ϵ that communicates at most γN bits. Fix $b = O(1)$. Then there exists a protocol that communicates at most $2\gamma N$ bits and accepts with probability $\widetilde{p}_{xy}$ on input $(x, y) \in \{\pm 1\}^N \times \{\pm 1\}^N$ such that:*

$$\mathbb{E}_{(x,y)\sim\xi_{-\sqrt{b/N}}}\left[\widetilde{p}_{xy}\right] \in \left[\frac{1-\epsilon'}{4N}, \frac{1+\epsilon'}{4N}\right], \quad \mathbb{E}_{(x,y)\sim\xi_0}\left[\widetilde{p}_{xy}\right] \in \left[\frac{(b+1)\beta - \epsilon'}{4N}, \frac{(b+1)\beta + \epsilon'}{4N}\right],$$

$$\mathbb{E}_{(x,y)\sim\xi_{\sqrt{b/N}}}\left[\widetilde{p}_{xy}\right] \in \left[\frac{(4b+1)\beta - \epsilon'}{4N}, \frac{(4b+1)\beta + \epsilon'}{4N}\right],$$

where $\epsilon' = e^{Cb}\epsilon$ for some universal constant C, and $\beta \in [1 - o(1), 1]$.

Proof. We first show that, for any $N = 2^n$, the existence of a Fourier sampler with the assumed parameters implies the existence of a protocol that communicates γN bits and accepts with probability $\widetilde{p}_{xy}$ for each pair (x, y), such that

$$\mathbb{E}_{(x,y)\sim\xi_0}\left[\left|\widetilde{p}_{xy} - \left(\frac{\langle x, y\rangle}{N}\right)^2\right|\right] \le \frac{\epsilon}{N}. \tag{6}$$

To see this, note that by the definition of Distributed Fourier Sampling and the assumed parameters for the protocol, there exists a protocol that communicates γN bits and samples from some distribution $\widetilde{p}_{xy}$ on bit-strings $s \in \{0,1\}^n$ such that

$$\epsilon \ge \mathbb{E}_{(x,y)\sim\xi_0}\left[\sum_{s\in\{0,1\}^n} |\widetilde{p}_{xy}(s) - p_{xy}(s)|\right] = \sum_{s\in\{0,1\}^n} \mathbb{E}_{x,y}[|\widetilde{p}_{xy}(s) - p_{xy}(s)|],$$

where $p_{xy}(s) = \left(\frac{1}{N}\sum_{z\in\{0,1\}^n}(-1)^{s\cdot z}x_z y_z\right)^2$; observe that $p_{xy}(0^n) = (\langle x, y\rangle/N)^2$. So there must exist s such that $\mathbb{E}_{(x,y)\sim\xi_0}[|\widetilde{p}_{xy}(s) - p_{xy}(s)|] \le \epsilon/N$. This implies the existence of a protocol which accepts with probability $\widetilde{p}_{xy}$ on input (x, y) such that $\mathbb{E}_{(x,y)\sim\xi_0}[|\widetilde{p}_{xy} - p_{xy}(s)|] \le \epsilon/N$. We can assume that $s = 0^n$ by replacing x with the string $x^{(s)}$ defined by $x_z^{(s)} = (-1)^{s\cdot z}x_z$, as $p_{xy}(s) = p_{x^{(s)}y}(0^n)$. As (x, y) were uniformly distributed, so are $(x^{(s)}, y)$.

Alice and Bob will apply their protocol which achieves the bound (6) to a pair of inputs $x', y' \in \{\pm 1\}^{2N}$ given by a shifted version of their original inputs (x, y). For any $p \in [-1, 1]$ and any $q \in [0, 1]$, given a pair of inputs $(x, y) \sim \xi_p$, Alice and Bob can produce a pair of inputs distributed according to ξ_{p+q-pq} by using their shared randomness to do the following: for each i, with probability q, replace the pair $(x_i, y_i) \in \{\pm 1\}^2$ with an equal pair picked uniformly at random from $\{\pm 1\}$. Then each of x_i and y_i is individually uniform, and

$$\Pr[x_i = y_i] = q + (1-q)\frac{1+p}{2} = \frac{1}{2} + \frac{p+q-pq}{2}.$$

Taking $p = \sqrt{b/N}$, $q = p/(1+p)$, this maps $\xi_{-p} \mapsto \xi_0$, $\xi_0 \mapsto \xi_{p/(1+p)}$, $\xi_p \mapsto \xi_{2p/(1+p)}$.

Alice and Bob then form a pair of inputs (x', y') by concatenating the original inputs with a pair of inputs in $\{\pm 1\}^N$ that differ at exactly $N/2$ positions, and applying the same random permutation of indices to each of their resulting strings. Call the resulting distribution on pairs (x', y'), when applied to a starting distribution ξ_r, ξ_r'. Then clearly

$$\Pr_{(x',y')\sim\xi_r'}[\langle x', y'\rangle = \Delta] = \Pr_{(x,y)\sim\xi_r}[\langle x, y\rangle = \Delta].$$

13

By Fact 8, there is a universal constant C such that for any $r \in [-0.01, 0.01]$,

$$\Pr_{(x',y') \sim \xi'_r}[\langle x', y' \rangle = \Delta] \leq C e^{2r^2 N} \Pr_{(x',y') \sim \xi_0^{2N}}[\langle x', y' \rangle = \Delta].$$

For $r \in \{0, p/(1+p), 2p/(1+p)\}$ and sufficiently large N, $0 \leq r \leq 3\sqrt{b/N}$. Write $\mathrm{Err}(x', y') := |\widetilde{p}_{x'y'} - (\langle x', y' \rangle / (2N))^2|$ for the expected inaccuracy when Alice and Bob apply their protocol to (x', y'). Then we have

$$
\begin{aligned}
\mathbb{E}_{(x',y') \sim \xi'_r}\left[\mathrm{Err}(x', y')\right] &= \sum_{\Delta=-N}^{N} \Pr_{(x',y') \sim \xi'_r}[\langle x', y' \rangle = \Delta]\mathbb{E}_{\langle x',y' \rangle = \Delta}\left[\mathrm{Err}(x', y')\right] \\
&\leq \sum_{\Delta=-N}^{N} e^{C'b} \Pr_{(x',y') \sim \xi_0^{2N}}[\langle x', y' \rangle = \Delta]\mathbb{E}_{\langle x',y' \rangle = \Delta}\left[\mathrm{Err}(x', y')\right] \\
&\leq e^{C'b}\mathbb{E}_{(x',y') \sim \xi_0^{2N}}\left[\mathrm{Err}(x', y')\right] \\
&\leq e^{C'b}\frac{\epsilon}{N},
\end{aligned}
$$

where the inner expectations are uniform over strings x', y' such that $\langle x', y' \rangle = \Delta$, C' is some other universal constant, and the last inequality follows from (6). By Fact 7, we have

$$\mathbb{E}_{(x',y') \sim \xi'_0}\left[\left(\frac{\langle x', y' \rangle}{2N}\right)^2\right] = \frac{1}{4N}, \quad \mathbb{E}_{(x',y') \sim \xi'_{p/(1+p)}}\left[\left(\frac{\langle x', y' \rangle}{2N}\right)^2\right] = \frac{1}{4}\left(\frac{1}{N} + \left(1 - \frac{1}{N}\right)\left(\frac{p}{1+p}\right)^2\right),$$

$$\mathbb{E}_{(x',y') \sim \xi'_{2p/(1+p)}}\left[\left(\frac{\langle x', y' \rangle}{N}\right)^2\right] = \frac{1}{4}\left(\frac{1}{N} + \left(1 - \frac{1}{N}\right)\left(\frac{2p}{1+p}\right)^2\right).$$

The lemma follows by inserting the definition of p. $\qquad\qquad\square$

5 Proof of Lemma 6: skewed anticoncentration bound

The proof of Lemma 6 is analogous to that of Lemma 2.5 in [16]. Indeed, although it is a generalisation of this result, somewhat remarkably the same proof technique goes through. The starting point is to prove an equivalent result for Gaussian measure.

Let γ^N denote the Gaussian distribution on $\mathbb{R}^N$ with density $(2\pi)^{-N/2} e^{\|x\|^2/2}$, and let Ξ_p denote the distribution on pairs (x, y) formed by choosing $x, z \sim \gamma^N$, and setting $y = px + \sqrt{1 - p^2}z$. Note that Ξ_p is the "Gaussian analogue" of the distribution ξ_p previously introduced.

Let $D(\cdot \| \cdot)$ denote the relative entropy,

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

and write $D_\gamma(X) := D(P \| \gamma)$, where X is a random variable with distribution P. We will need the following technical claim:

Fact 9 (Chakrabarti and Regev [16]). *For all $\epsilon, \alpha_0 > 0$ there exists a $\delta > 0$ such that for any probability distribution P on $\mathbb{R}$ satisfying $D_\gamma(P) < \delta$, any $z \in \mathbb{R}$, and any $0 < \alpha \leq \alpha_0$, we have*

$$\mathbb{E}_{x \sim P}[\cosh(\alpha x + z)] \geq e^{\alpha^2/2} - \epsilon.$$

14

We will prove the following result:

Theorem 10. *For all $c, \epsilon > 0$ there exists $\delta > 0$ such that for large enough N and $0 \leq \eta \leq c/\sqrt{N}$, all $A, B \subseteq \mathbb{R}^N$ such that $\gamma^N(A), \gamma^N(B) \geq 2^{-\delta N}$, and all s,*

$$\frac{1}{2}\left(e^s \Pr_{(x,y) \sim \Xi_{-\eta}}[x \in A \wedge y \in B] + e^{-s} \Pr_{(x,y) \sim \Xi_{\eta}}[x \in A \wedge y \in B]\right) \geq (1-\epsilon)\gamma^N(A)\gamma^N(B).$$

Chakrabarti and Regev's Theorem 3.5 is Theorem 10 with $s = 0$ [16]. The proof of Theorem 10 is essentially identical to this special case. We therefore do not include all the details, and refer to [16] where possible. In the proof, $\beta_1, \ldots, \beta_6$ are constants that can be taken to be arbitrarily small.

Proof of Theorem 10. Write, for any $-1 \leq \zeta \leq 1$,

$$p_\zeta := \Pr_{(x,y) \sim \Xi_\zeta}[x \in A \wedge y \in B].$$

It is shown in [16, Proof of Theorem 3.5] that there exist subsets $A' \subseteq A$, $B' \subseteq B$ such that $\gamma^N(A') \geq (1 - \beta_2)\gamma^N(A)$, $\gamma^N(B') \geq (1 - \beta_2)\gamma^N(B)$ and

$$p_\zeta \geq (1 - \zeta^2)^{-N/2} e^{-\zeta^2(1+\beta_1)N/(1-\zeta^2)} \mathbb{E}_{x \sim \gamma^N|_{A'}, y \sim \gamma^N|_{B'}}[e^{\zeta\langle x,y\rangle/(1-\zeta^2)}]\gamma^N(A')\gamma^N(B').$$

Applying this inequality with $\zeta = \pm\eta$, we obtain that the quantity we wish to bound, which we can write as $\frac{1}{2}(e^s p_{-\eta} + e^{-s}p_\eta)$, is lower-bounded by

$$(1-\eta^2)^{-N/2}e^{-\eta^2(1+\beta_1)N/(1-\eta^2)}\mathbb{E}_{x \sim \gamma^N|_{A'}, y \sim \gamma^N|_{B'}}\left[\frac{1}{2}e^s e^{-\eta\langle x,y\rangle/(1-\eta^2)} + \frac{1}{2}e^{-s}e^{\eta\langle x,y\rangle/(1-\eta^2)}\right]\gamma^N(A')\gamma^N(B'). \tag{7}$$

We next observe that

$$\frac{1}{2}e^s e^{-\eta\langle x,y\rangle/(1-\eta^2)} + \frac{1}{2}e^{-s}e^{\eta\langle x,y\rangle/(1-\eta^2)} = \cosh(\eta\langle x, y\rangle/(1 - \eta^2) - s).$$

Let $B'' \subseteq B'$ be the set of all $y \in B'$ for which

$$\mathbb{E}_{x \sim \gamma^N|_{A'}}[\cosh(\eta\langle x, y\rangle/(1 - \eta^2) - s)] \leq (1 - \beta_3)e^{(\eta/(1-\eta^2))^2(1-\beta_1)N/2}.$$

We will show that this set is relatively small, $\gamma^N(B'') \leq \beta_4 \gamma^N(B')$, implying that

$$\mathbb{E}_{x \sim \gamma^N|_{A'}, y \sim \gamma^N|_{B'}}[\cosh(\eta\langle x, y\rangle/(1 - \eta^2) - s)] \geq (1 - \beta_4)(1 - \beta_3)e^{(\eta/(1-\eta^2))^2(1-\beta_1)N/2}$$

and hence

$$
\begin{aligned}
(7) \quad &\geq \quad (1 - \eta^2)^{-N/2}e^{-\eta^2(1+\beta_1)N/(1-\eta^2)}(1 - \beta_4)(1 - \beta_3)e^{(\eta/(1-\eta^2))^2(1-\beta_1)N/2}\gamma^N(A')\gamma^N(B') \\
&\geq \quad e^{N\eta^2/2}e^{-\eta^2(1+\beta_1)N/(1-\eta^2)}(1 - \beta_4)(1 - \beta_3)e^{(\eta/(1-\eta^2))^2(1-\beta_1)N/2}(1 - \beta_2)^2\gamma^N(A)\gamma^N(B) \\
&\geq \quad (1 - \epsilon)\gamma^N(A)\gamma^N(B)
\end{aligned}
$$

for sufficiently small $\beta_1, \beta_2, \beta_3, \beta_4$ and sufficiently large N. Assume the contrary, i.e. that $\gamma^N(B'') > \beta_4 \gamma^N(B') > \beta_4(1-\beta_2)e^{-\delta N}$. Then it is shown in [16, Proof of Theorem 3.5] that there exists $y \in B''$ such that the distribution of $\langle x, y\rangle$, where $x \sim \gamma^N|_{A'}$, is given by the distribution of $\alpha r X + rY$ for

15

some $\alpha \in [1 - \beta_6, 1]$; some $r \in [\sqrt{(1 - \beta_1)N}, \sqrt{(1 + \beta_1)N}]$; and random variables X and Y such that

$$\Pr_Y[D_\gamma(X|Y) \leq \beta_5] \geq 1 - \beta_5;$$

here β_5 and β_6 are arbitrarily small constants (assuming that δ and the other β_i constants are sufficiently small). By Fact 9, for sufficiently small β_7,

$$
\begin{aligned}
\mathbb{E}_{x \sim \gamma^N|_{A'}}[\cosh(\eta\langle x, y\rangle/(1 - \eta^2) - s)] &= \mathbb{E}[\cosh(\eta(\alpha r X + rY)/(1 - \eta^2) - s)] \\
&\geq (1 - \beta_5)(e^{(\eta/(1-\eta^2)\alpha r)^2/2} - \beta_7) \\
&\geq (1 - \beta_5)(e^{(\eta/(1-\eta^2))^2(1-\beta_6)^2(1-\beta_1)N/2} - \beta_7) \\
&> (1 - \beta_3)e^{(\eta/(1-\eta^2))^2(1-\beta_1)N/2},
\end{aligned}
$$

contradicting the assumption that $y \in B''$. $\qquad\square$

We can now prove the corollary we need for the boolean cube (which is Lemma 6 stated in a slightly more general form).

Corollary 11. *For all $c, \epsilon > 0$ there exists $\delta > 0$ such that for large enough N and $0 \leq p \leq c/\sqrt{N}$, all rectangles $R \subseteq \{\pm 1\}^N \times \{\pm 1\}^N$ such that $\xi_0(R) \geq 2^{-\delta N}$, and all s,*

$$\frac{1}{2}\left(e^s\xi_{-p}(R) + e^{-s}\xi_p(R)\right) \geq (1 - \epsilon)\xi_0(R).$$

Proof. Again, the argument is essentially the same as in [16]. For $R = A \times B$, let $A' = \{x \in \mathbb{R}^N : \text{sgn}(x) \in A\}$, and similarly for B'. Then one can check that, for any $\eta \in [-1, 1]$,

$$\Pr_{(x,y)\sim\Xi_\eta}[x \in A' \wedge y \in B'] = \xi_p(R)$$

for $p = 1 - \frac{2}{\pi}\arccos\eta$. In particular, taking $\eta = 0$,

$$\min\{\gamma^N(A'), \gamma^N(B')\} \geq \gamma^N(A')\gamma^N(B') = \Pr_{(x,y)\sim\Xi_0}[x \in A' \wedge y \in B'] = \xi_0(R) \geq 2^{-\delta N}.$$

For $p \leq c/\sqrt{N}$ and sufficiently large N, $\eta \leq c'/\sqrt{N}$ for some constant c', so we can apply Theorem 10 to infer the claim. $\qquad\square$

6 Proofs of further technical facts

In this section we prove some claims about the distributions ξ_p, ξ_p' used in the proof of Lemma 5. Recall that ξ_p is the distribution on pairs $(x, y) \in \{\pm 1\}^N$ defined in Definition 1.

Fact 7 (restated).

$$\mathbb{E}_{(x,y)\sim\xi_p}\left[\left(\frac{\langle x, y\rangle}{N}\right)^2\right] = \frac{1}{N} + \left(1 - \frac{1}{N}\right)p^2.$$

Proof.

$$\mathbb{E}_{(x,y)\sim\xi_p}\left[\left(\frac{\langle x,y\rangle}{N}\right)^2\right] = \frac{1}{N^2}\mathbb{E}_{(x,y)\sim\xi_p}\left[\left(\sum_i x_i y_i\right)^2\right]$$

$$= \frac{1}{N} + \frac{1}{N^2}\sum_{i\neq j}\mathbb{E}_{(x,y)\sim\xi_p}[x_i y_i x_j y_j]$$

$$= \frac{1}{N} + \frac{1}{N^2}\sum_{i\neq j}\mathbb{E}_{(x,y)\sim\xi_p}[x_i y_i]\mathbb{E}_{(x,y)\sim\xi_p}[x_j y_j]$$

$$= \frac{1}{N} + \frac{N(N-1)}{N^2}(\mathbb{E}_{(x,y)\sim\xi_p}[x_i y_i])^2$$

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

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

$\square$

Fact 8 (restated). *Let ξ'_p be the distribution on pairs $(x', y') \in \{\pm 1\}^{2N}$ formed by choosing $(x,y) \sim \xi_p$, then setting $x' = \sigma(x, x'')$, $y' = \sigma(y, y'')$, where $x'', y'' \in \{\pm 1\}^N$ are fixed strings such that $\langle x'', y''\rangle = 0$, and $\sigma \in S_{2N}$ is a random permutation of the indices of the strings. Then there exists a universal constant C such that, for sufficiently large N and any Δ, and $p \in [-0.01, 0.01]$,*

$$\Pr_{(x',y')\sim\xi'_p}[\langle x', y'\rangle = \Delta] \leq Ce^{p^2 N}\Pr_{(x',y')\sim\xi_0^{2N}}[\langle x', y'\rangle = \Delta].$$

Proof. By the definition of the distributions ξ, ξ', we have

$$\Pr_{(x',y')\sim\xi'_p}[\langle x', y'\rangle = \Delta] = \left(\frac{1+p}{2}\right)^{(N+\Delta)/2}\left(\frac{1-p}{2}\right)^{(N-\Delta)/2}\binom{N}{\frac{N+\Delta}{2}},$$

$$\Pr_{(x',y')\sim\xi_0^{2N}}[\langle x', y'\rangle = \Delta] = 2^{-2N}\binom{2N}{N+\frac{\Delta}{2}}.$$

Setting $k = (N+\Delta)/2$, the ratio R of these two quantities is thus

$$R = 2^N(1+p)^k(1-p)^{N-k}\frac{\binom{N}{k}}{\binom{2N}{k+N/2}} \leq 2^N e^{p(2k-N)}\frac{\binom{N}{k}}{\binom{2N}{k+N/2}}.$$

We now upper-bound this expression using the binomial coefficient bounds, valid for all integer $k \in [1, N-1]$,

$$\sqrt{\frac{N}{8k(N-k)}}e^{Nh(k/N)} \leq \binom{N}{k} \leq \sqrt{\frac{N}{2\pi k(N-k)}}e^{Nh(k/N)}, \tag{8}$$

where $h(x) := -x\ln x - (1-x)\ln(1-x)$ is the binary entropy measured in nats [31]. Then for some universal constant C,

$$R \leq C2^N e^{p(2k-N)}\sqrt{\frac{(k+N/2)(3N/2-k)}{k(N-k)}}e^{N(h(k/N)-2h(k/(2N)+1/4))}.$$

First consider $k \in [0, 0.1N] \cup [0.9N, N]$. For these values of k, using $k \le N$ and (8), there is a universal constant C such that

$$2^N e^{p(2k-N)} \frac{\binom{N}{k}}{\binom{2N}{k+N/2}} \le 2^N e^{Np} \frac{\binom{N}{0.1N}}{\binom{2N}{N/2}} \le C 2^N e^{Np} e^{Nh(0.1)-2Nh(1/4)} \le C e^{N(p-0.1)},$$

implying that, for $p \le 0.01$, $R \le 1$ for sufficiently large N.

Now consider the range $k \in [0.1N, 0.9N]$. Defining $x = k/N$ and combining terms, we have

$$R \le C \sqrt{\frac{(x+1/2)(3/2-x)}{x(1-x)}} e^{N(\ln 2 + p(2x-1)+h(x)-2h(x/2+1/4))}.$$

For these values of x,
$$R \le C' e^{N(\ln 2 + p(2x-1)+h(x)-2h(x/2+1/4))}$$

for some universal constant C', so to prove the bound claimed in the lemma it suffices to show that $f(x) := \ln 2 + p(2x - 1) + h(x) - 2h(x/2 + 1/4) \le p^2$ for all $x \in [0, 1]$. By concavity of h, $\ln 2 + h(x) = h(1/2) + h(x) \le 2h(x/2 + 1/4)$, so $f(x) \le p(2x - 1)$, which is negative for $x < 1/2$. So, defining $F(x) := f(x + 1/2) - 2px = \ln 2 + h(x + 1/2) - 2h(x/2 + 1/2)$, we want to find an upper bound on $F(x)$ over $x \in [0, 1/2]$. We will show that $F(x) \le -x^2$ and hence $f(x + 1/2) \le 2px - x^2 \le p^2$, where we take the maximum over x.

We have $F(0) = 0$, and

$$F'(x) := \frac{d}{dx} F(x) = \ln \left(\frac{1 - 2x}{1 + 2x} \right) - \ln \left(\frac{1 - x}{1 + x} \right),$$

so $F'(0) = 0$ as well. Thus it is sufficient to show that $F''(x) \le -2$ for all $x \in [0, 1/2]$ to prove the desired bound. One can calculate that

$$F''(x) = -\frac{2 + 4x^2}{1 - 5x^2 + 4x^4} < -2$$

for x in this range, completing the proof. $\qquad\square$

Acknowledgements

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 $\langle\psi|M|\psi\rangle \ge 2/3$ or $\langle\psi|M|\psi\rangle \le 1/3$; 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 [27]. This protocol was proposed by Raz [36] (and subsequently also described by Klartag and Regev [25]), 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 12 (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 $|\phi_1\rangle, \ldots, |\phi_K\rangle$, 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

$$\Pr_{|\psi^\perp\rangle}\left[\langle\phi_i|M|\phi_i\rangle \geq \frac{1}{2}\right]. \tag{9}$$

We can expand

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

and using $\langle\psi|M|\psi\rangle \leq 1/3$ and a union bound, can upper-bound (9) as

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

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

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

using Lemma 12 with $r=1$, $\delta=(\epsilon/24)^2(N-1)-1$, and assuming that $\epsilon\gg N^{-1/2}$. For the second term, we use

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

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

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

together with Lemma 12 and the bound

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

to obtain

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

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 $0\leq x\leq 1$, we have

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

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 [38]. 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 $0<D<N^{1/2}$,

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

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

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

In this appendix we give a simple direct proof of Corollary 2 in the special case of deterministic query algorithms; that is, algorithms that choose which bits to query deterministically, then sample from some distribution depending on the bits queried.

Proposition 13. *For sufficiently small constant $\epsilon > 0$, any deterministic classical query algorithm that solves Fourier Sampling on N input bits must make $\Omega(N)$ 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 $\|\widetilde{p}_h - p_h\|_1 \leq \epsilon$, 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 Lemma 5, 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

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

where x is uniformly random. Because the uniform distribution is permutation-invariant, 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 $\left(\sum_i y_i \right)^2$ (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 $\mathbb{E}_y \left[| \sum_i y_i | \right] = O(\sqrt{k})$, $\mathbb{E}_z \left[| \sum_i z_i | \right] = O(\sqrt{m})$. 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

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

21

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. $\square$

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] S. Brierley, A. Kosowski, M. Markiewicz, T. Paterek, and A. Przysiezna. Non-classicality of temporal correlations. *Phys. Rev. Lett.*, 115:120404, 2015. `arXiv:1501.03505`.

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

[12] 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`.

[13] 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`.

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

[15] 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`.

[16] 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`.

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

[18] 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`.

[19] 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`.

[20] 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`.

[21] 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.

[22] 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`.

[23] 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.

[24] R. Jain and H. Klauck. The partition bound for classical communication complexity and query complexity. In *Proc. 25th Annual IEEE Conf. Computational Complexity*, pages 247–258, 2010. `arXiv:0910.4266`.

[25] 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`.

[26] 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`.

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

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

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

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

[31] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes*. North-Holland, Amsterdam, 1983.

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

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

[34] 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`.

[35] 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`.

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

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

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

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

[40] 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.

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