Quantum Versus Randomized Communication Complexity, with Efficient Players

Uma Girish
Department of Computer Science, Princeton University, NJ, USA
ugirish@cs.princeton.edu

Ran Raz
Department of Computer Science, Princeton University, NJ, USA
ranr@cs.princeton.edu

Avishay Tal
Department of Electrical Engineering and Computer Sciences,
University of California at Berkeley, CA, USA
atal@berkeley.edu

Abstract

We study a new type of separations between quantum and classical communication complexity, separations that are obtained using quantum protocols where all parties are **efficient**, in the sense that they can be implemented by small quantum circuits, with oracle access to their inputs. Our main result qualitatively matches the strongest known separation between quantum and classical communication complexity [8] and is obtained using a quantum protocol where all parties are efficient. More precisely, we give an explicit partial Boolean function f over inputs of length N, such that:

(1) f can be computed by a simultaneous-message quantum protocol with communication complexity $\text{polylog}(N)$ (where at the beginning of the protocol Alice and Bob also have $\text{polylog}(N)$ entangled EPR pairs).

(2) Any classical randomized protocol for f, with any number of rounds, has communication complexity at least $\tilde{\Omega}\left(N^{1/4}\right)$.

(3) All parties in the quantum protocol of Item (1) (Alice, Bob and the referee) can be implemented by quantum circuits of size $\text{polylog}(N)$ (where Alice and Bob have oracle access to their inputs).

Items (1), (2) qualitatively match the strongest known separation between quantum and classical communication complexity, proved by Gavinsky [8]. Item (3) is new. (Our result is incomparable to the one of Gavinsky. While he obtained a quantitatively better lower bound of $\Omega\left(N^{1/2}\right)$ in the classical case, the referee in his quantum protocol is inefficient).

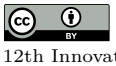
Exponential separations of quantum and classical communication complexity have been studied in numerous previous works, but to the best of our knowledge the efficiency of the parties in the quantum protocol has not been addressed, and in most previous separations the quantum parties seem to be inefficient. The only separations that we know of that have efficient quantum parties are the recent separations that are based on lifting [10, 5]. However, these separations seem to require quantum protocols with at least two rounds of communication, so they imply a separation of two-way quantum and classical communication complexity but they do not give the stronger separations of simultaneous-message quantum communication complexity vs. two-way classical communication complexity (or even one-way quantum communication complexity vs. two-way classical communication complexity).

Our proof technique is completely new, in the context of communication complexity, and is based on techniques from [15]. Our function f is based on a lift of the FORRELATION problem, using XOR as a gadget.

2012 ACM Subject Classification Theory of computation → Communication complexity; Theory of computation → Quantum complexity theory

Keywords and phrases Exponential Separation, Quantum, Randomized, Communication, Complexity, Forrelation

Digital Object Identifier 10.4230/LIPIcs.ITCS.2021.54



Related Version A full version of the paper is available at https://arxiv.org/abs/1911.02218.

Funding *Uma Girish*: Research supported by the Simons Collaboration on Algorithms and Geometry, by a Simons Investigator Award and by the National Science Foundation grant No. CCF-1714779.
Ran Raz: Research supported by the Simons Collaboration on Algorithms and Geometry, by a Simons Investigator Award and by the National Science Foundation grant No. CCF-1714779.
Avishay Tal: Part of this work was done when the author was a postdoc at the Department of Computer Science, Stanford University. Partially supported by a Motwani Postdoctoral Fellowship and by NSF grant CCF-1763311.

1 Introduction

Exponential separations between quantum and classical communication complexity have been established in various models and settings. These separations give explicit examples of partial functions that can be computed by quantum protocols with very small communication complexity, while any classical randomized protocol requires significantly higher communication complexity. However, to the best of our knowledge, in all these works the efficiency of the quantum players in the quantum protocol has not been addressed and in most of these separations, the quantum players are inefficient.

Communication complexity studies the amount of communication needed to perform computational tasks that depend on two (or more) inputs, each given to a different player. The efficiency of the players in a communication complexity protocol is usually not addressed. If the players need to read their entire inputs, their time complexity is at least the length of the inputs. However, the inputs may be represented compactly by a black box and (particularly in the quantum case) we can hope for players that can be implemented very efficiently by small (say, poly-logarithmic size) quantum circuits, with oracle access to their inputs.

Our main result qualitatively matches the strongest known separation between quantum and classical communication complexity [8] and is obtained using quantum protocols where all players are efficient. To prove our results we use a completely different set of techniques, based on techniques from the recent oracle separation of BQP and PH [15].

1.1 Previous Work

The relative power of quantum and classical communication complexity has been studied in numerous of works. While it is unknown whether quantum communication can offer exponential advantage over randomized communication for total functions, a series of works gave explicit examples of partial Boolean functions (promise problems) that have quantum protocols with very small communication complexity, while any classical protocol requires exponentially higher communication complexity. The history of exponential advantage of quantum communication, that is most relevant to our work, is briefly summarized below.

Buhrman, Cleve and Wigderson gave the first (exponential) separation between zero-error quantum communication complexity and classical deterministic communication complexity [4]. Raz gave the first exponential separation between two-way quantum communication complexity and two-way randomized communication complexity [14]. Bar-Yossef et al [3] (for search problems) and Gavinsky et al [9] (for promise problems) gave the first (exponential) separations between one-way quantum communication complexity and one-way randomized communication complexity. Klartag and Regev gave the first (exponential) separation between one-way quantum communication complexity and two-way random-

ized communication complexity [16]. Finally, Gavinsky gave an (exponential) separation between simultaneous-message quantum communication complexity and two-way randomized communication complexity [8].

We note that Gavinsky's work is the strongest separation known today and essentially subsumes the separations discussed above. More precisely, Gavinsky [8] gave an explicit partial Boolean function f over inputs of length N, such that:

1. f can be computed by a simultaneous-message quantum protocol with communication complexity polylog(N): Alice and Bob simultaneously send quantum messages of length polylog(N) to a referee, who performs a quantum measurement on the messages and announces the answer. (At the beginning of the protocol Alice and Bob also have polylog(N) entangled EPR pairs).
   We note that this also implies a one-way quantum protocol where Alice sends a message of length polylog(N) qubits to Bob, who performs a measurement and announces the answer (or vice versa).
2. Any classical randomized protocol for f has communication complexity at least $\Omega\left(N^{1/2}\right)$.

A drawback of Gavinsky's separation, in the context of our work, is that the referee in his quantum protocol is inefficient as it is required to perform $O(N)$ quantum operations (and this seems to be crucial in his lower bound proof).

As mentioned before, to the best of our knowledge, the efficiency of the quantum players has not been addressed in previous works on separations of quantum and classical communication complexity. The only separations that we know of that do have efficient quantum parties are the separations that follow from the recent randomized query-to-communication lifting theorems of [10, 5], applied to problems for which we know that quantum decision trees offer an exponential advantage over randomized ones, such as the FORRELATION problem of [1, 2]. However, lifting with the gadgets used in [10, 5] seems to require quantum protocols with two rounds of communication. Thus, these theorems only imply a separation of two-way quantum and classical communication complexity and do not give the stronger separations of simultaneous-message quantum communication complexity vs. two-way classical communication complexity (or even one-way quantum communication complexity vs. two-way classical communication complexity).

1.2 Our Result

We recover Gavinsky's state of the art separation, using entirely different techniques. While the parameters in our bounds are weaker, our quantum protocol is *efficient*, in the sense that it involves just polylog(N) amount of work by Alice, Bob and the referee, when the players have blackbox access to their inputs. In other words, the output of the entire simultaneous protocol can be described by a polylog(N) size quantum circuit, with oracle access to the inputs.

More precisely, our main result gives an explicit partial Boolean function f over inputs of length N, such that:

1. As in Gavinsky's work, f can be computed by a simultaneous-message quantum protocol with communication complexity polylog(N): Alice and Bob simultaneously send quantum messages of length polylog(N) to a referee, who performs a quantum measurement on the messages and announces the answer. (At the beginning of the protocol Alice and Bob also have polylog(N) entangled EPR pairs).
   As before, this also implies a one-way quantum protocol where Alice sends a message of length polylog(N) qubits to Bob, who performs a measurement and announces the answer (or vice versa).

2. Any classical randomized protocol for f has communication complexity at least $\tilde{\Omega}\left(N^{1/4}\right)$.
3. All parties in the quantum protocol of Item (1) (Alice, Bob and the referee) can be implemented by quantum circuits of size polylog(N) (where Alice and Bob have oracle access to their input).

The problem that we define is a lift of the FORRELATION problem of [1, 2, 15] with XOR as the gadget. Our proof technique follows the Fourier-analysis framework of [15]. Our proof offers an entirely new and possibly simpler approach for communication complexity lower bounds. We believe this technique may be applicable in a broader setting. We note that lower bounds for lifting by XOR, using a Fourier-analysis approach, were previously studied in [13, 11].

1.3 Our Communication Complexity Problem

Let $N = 2^n$ and H_N be the $N \times N$ normalized Hadamard matrix. Let $x = (x_1, x_2)$ be an input where $x_1, x_2 \in \{-1, 1\}^N$. We define the forrelation of x as the correlation between the second half x_2 and the Hadamard transform of the first half x_1.

$$forr(x) := \left\langle \frac{1}{\sqrt{N}} H_N(x_1) \middle| \frac{1}{\sqrt{N}} x_2 \right\rangle$$

The communication problem for which our separation holds is a lift of the forrelation problem of [15], with XOR as the gadget. Let $x, y \in \{-1, 1\}^{2N}$. Alice gets x and Bob gets y and their goal is to compute the partial function F defined by

$$F(x, y) := \begin{cases} 1 & \text{if } forr(x \cdot y) \geq \frac{1}{200} \cdot \frac{1}{\ln N} \\ -1 & \text{if } forr(x \cdot y) \leq \frac{1}{400} \cdot \frac{1}{\ln N}. \end{cases}$$

Here $x \cdot y$ refers to the coordinate-wise product of the vectors x, y. We refer to this problem as the forrelation problem.

▶ **Theorem 1.** *The forrelation problem can be solved in the quantum simultaneous with entanglement model with $O(\log^3 N)$ bits of communication, when Alice and Bob are given access to $O(\log^3 N)$ bits of shared entanglement. Moreover, the protocol is efficient, as it can be implemented by a $O(\log^3 N)$ size quantum circuit with oracle access to inputs.*

The quantum upper bound on F follows from the fact that the XOR of the inputs can be computed by a simultaneous-message quantum protocol, when the players share entanglement, and the fact that $forr(x)$ can be estimated by a small size quantum circuit [1, 2, 15].

▶ **Theorem 2.** *The randomized bounded-error interactive communication cost of the forrelation problem is $\tilde{\Omega}(N^{\frac{1}{4}})$.*

1.4 An Overview of the Lower Bound

In this section, we outline the proof of the lower bound. We use the forrelation distribution $\mathcal{D}$ on $\{-1, 1\}^{2N}$ as defined by [15]. We define a distribution $\mathcal{V}$ on inputs to the communication problem, obtained by sampling $z \sim \mathcal{D}$, and $x \in \{-1, 1\}^{2N}$ uniformly at random, and setting $y := x \cdot z$. Alice gets x and Bob gets y. It can be shown that the distribution $\mathcal{V}$ has considerable support over the yes instances of F, while the uniform distribution $\mathcal{U}$ on $\{-1, 1\}^{4N}$ has large support over the no instances of F. This fact along with the following theorem implies a lower bound on the randomized communication cost of F.

▶ **Theorem 3.** *Consider the following distribution. A string $z \in \{-1,1\}^{2N}$ is drawn from the forrelation distribution, $x \sim U_{2N}$ is drawn uniformly and $y := x \cdot z$. Alice gets x and Bob gets y. Given any deterministic communication protocol $C : \{-1,1\}^{2N} \times \{-1,1\}^{2N} \to \{-1,1\}$ of cost $c \geq 1$, its expectation when the inputs are drawn from this distribution is close to when the inputs are drawn from the uniform distribution. That is,*

$$\left| \mathop{\mathbb{E}}_{\substack{x \sim U_{2N} \\ z \sim \mathcal{D}}} [C(x, x \cdot z)] - \mathop{\mathbb{E}}_{x,y \sim U_{2N}} [C(x,y)] \right| \leq O\left(\frac{c^2}{N^{1/2}}\right).$$

In other words, no deterministic protocol of cost $o(N^{1/4})$ has considerable advantage in distinguishing the above distribution from the uniform distribution.

We now outline the proof of this theorem. Any cost c protocol induces a partition of the input space into at most 2^c rectangles. Let $A \times B$ be any rectangle, and let $\mathbb{1}_A, \mathbb{1}_B : \{-1,1\}^{2N} \to \{0,1\}$ be the indicator functions of A and B respectively. Note that for all distributions $\mathcal{S}$ on $\{-1,1\}^{2N}$, we have

$$\mathop{\mathbb{E}}_{z \sim \mathcal{S}, x \sim U_{2N}} [\mathbb{1}_A(x)\mathbb{1}_B(x \cdot z)] = \mathop{\mathbb{E}}_{z \sim \mathcal{S}} [(\mathbb{1}_A * \mathbb{1}_B)(z)].$$

Here, the notation $f * g$ refers to the convolution of Boolean functions f and g. This identity implies that our goal is to show that the expectation of the function $\sum_{A \times B}(\mathbb{1}_A * \mathbb{1}_B)(z)$ over a uniformly distributed z is close to the expectation over $z \sim \mathcal{D}$. An essential contribution of the works of [15] and [7] is the following result. For any family of functions $\mathcal{F}$ that is closed under restrictions, to show that the family is fooled by the forrelation distribution, it suffices to bound the ℓ_1-norm of the second level Fourier coefficients of the family. More precisely, the maximum advantage of a function $f \in \mathcal{F}$ in distinguishing the uniform distribution and $\mathcal{D}$, is at most $O\left(\frac{1}{\sqrt{N}}\right)$ times the maximum second level Fourier mass of a function $f \in \mathcal{F}$. Since small cost communication protocols form a family of functions closed under restrictions, the same reasoning applies here. In this paper however, we present a complete proof of this connection. We then provide the following bound on the second level Fourier mass corresponding to a small cost protocol.

▷ Claim 1. Let $C(x,y) : \{-1,1\}^{2N} \times \{-1,1\}^{2N} \to \{-1,1\}$ be any deterministic protocol of cost $c \geq 1$, let $D(x,z) : \mathbb{R}^{2N} \times \mathbb{R}^{2N} \to \mathbb{R}$ refer to the unique multilinear extension of $C(x, x \cdot z)$ and $H : \mathbb{R}^{2N} \to \mathbb{R}$ be defined by $H(z) = \mathbb{E}_{x \sim U_{2N}} D(x,z)$. Then,

$$L_2(H) \triangleq \sum_{|S|=2} |\widehat{H}(S)| \leq 120c^2.$$

We now describe the proof of this claim. Let $A \times B$ be a rectangle in the partition induced by the cost c protocol. An important property of the convolution of two functions f, g is that for all subsets $S \subseteq [n]$, we have $\widehat{f * g}(S) = \widehat{f}(S)\widehat{g}(S)$. This, along with Cauchy-Schwarz implies that

$$\sum_{|S|=2} \left|\widehat{\mathbb{1}_A * \mathbb{1}_B}(S)\right| = \sum_{|S|=2} \left|\widehat{\mathbb{1}_A}(S)\widehat{\mathbb{1}_B}(S)\right| \leq \left(\sum_{|S|=2} \widehat{\mathbb{1}_A}(S)^2\right)^{1/2} \left(\sum_{|S|=2} \widehat{\mathbb{1}_B}(S)^2\right)^{1/2}.$$

We then use a well known inequality on Fourier coefficients. It appears as "Level-k Inequalities" in Ryan Odonnell's book [12, Chapter 9.5] and it states that for a function $f : \{-1,1\}^n \to \{0,1\}$ with expectation $\mathbb{E}[f] = \alpha$, for any $k \leq 2\ln(1/\alpha)$, we have $\sum_{|S|=k} \left(\widehat{f}(S)\right)^2 \leq$

$O(\alpha^2 \ln^k(1/\alpha))$. For simplicity, assume that $|A| = |B| = 2^{(n-c)/2}$. The previous paragraphs and the assumption that $\mathbb{E}[\mathbb{1}_A], \mathbb{E}[\mathbb{1}_B] = \frac{1}{2^{c/2}}$ imply that the advantage of a rectangle is at most $O\left(\frac{1}{\sqrt{N}} \frac{1}{2^c} c^2\right)$. Adding the contributions from all rectangles implies that the advantage of a cost c protocol is at most $O\left(\frac{c^2}{\sqrt{N}}\right)$. This implies that every protocol of cost $o(N^{1/4})$ has advantage at most $o(1)$ in distinguishing between $\mathcal{U}$ and $\mathcal{V}$. The bound in the case of a general partition follows from a concavity argument. This completes the proof overview.

Open Questions

We conjecture that the correct randomized communication complexity for this problem is $\tilde{\Omega}(\sqrt{N})$ and that the above proof technique can be strengthened to show this. One way to do this would be to show a better bound on the Fourier coefficients of deterministic communication protocols. In particular, it would suffice to show a bound of $O(c \cdot \operatorname{poly}\log(N))$ on the second level Fourier mass of protocols with c-bits of communication.

2 Preliminaries

For $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, 2, \ldots, n\}$. For a vector $x \in \mathbb{R}^n$ and $i \in [n]$, we refer to the i-th coordinate of x by either $x(i)$ or x_i. For a subset $S \subset [n]$, let $x_S \in \mathbb{R}^{|S|}$ be the restriction of x to coordinates in S. For vectors $x, y \in \mathbb{R}^n$, let $x \cdot y$ be their point-wise product, i.e., the vector whose i-th coordinate is $x_i y_i$. Let $\langle x|y\rangle$ be the real inner product $\sum_i x_i y_i$ between x and y. Let v^{-1} be the coordinate-wise inverse of a vector $v \in (\mathbb{R} \setminus 0)^n$.

2.1 Fourier Analysis on the Boolean Hypercube

The set $\{-1, 1\}^n$ is referred to as the Boolean hypercube in n dimensions, or the n-dimensional hypercube. We sometimes refer to it by $\{0, 1\}^n$, using the bijection mapping $(x_1, \ldots, x_n) \in \{0, 1\}^n$ to $((-1)^{x_1}, \ldots, (-1)^{x_n}) \in \{-1, 1\}^n$. We also represent elements of $\{-1, 1\}^n$ by elements of $[2^n]$, using the bijection mapping $((-1)^{x_1}, \ldots, (-1)^{x_n}) \in \{-1, 1\}^n$ to $1 + \sum_{i=1}^n 2^{i-1} x_i \in [2^n]$. We typically use N to denote 2^n. Let $\mathbb{I}_n$ denote the $n \times n$ identity matrix. Let U_n be the uniform distribution on $\{-1, 1\}^n$. Let $\mathcal{F} := \{F : \{-1, 1\}^n \to \mathbb{R}\}$ be the set of all functions from the n-dimensional hypercube to the real numbers. This is a real vector space of dimension 2^n. We define an inner product over this space. For every $f, g, \in \mathcal{F}$, let

$$\langle f, g\rangle := \mathop{\mathbb{E}}_{x \sim U_n} [f(x) g(x)] .$$

For any universe $\mathcal{U}$ and a subset $S \subseteq \mathcal{U}$, we use $\mathbb{1}_S : \mathcal{U} \to \{0, 1\}$ to refer to the indicator function of S defined by:

$$\mathbb{1}_S(x) := \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{otherwise.} \end{cases}$$

The set of indicator functions of singleton sets $\{\mathbb{1}_{\{a\}} : a \in \{-1, 1\}^n\}$ is the standard orthogonal basis for $\mathcal{F}$. The character functions form an orthonormal basis for $\mathcal{F}$. These are functions $\chi_S : \{-1, 1\}^n \to \{-1, 1\}$ associated to every set $S \subseteq [n]$ and are defined at every point $x \in \{-1, 1\}^n$ by $\chi_S(x) := \prod_{i \in S} x_i$. For a function $f \in \mathcal{F}$, and $S \subseteq [n]$, we define its S-th Fourier coefficient to be $\widehat{f}(S) := \mathbb{E}_{x \sim U_n} [f(x) \chi_S(x)]$. Every $f \in \mathcal{F}$ can be expressed as $f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \chi_S(x)$. For $f : \{-1, 1\}^n \to \mathbb{R}$ and $k \in \{0, \ldots, n\}$, let $L_k(f) := \sum_{S \subseteq [n], |S| = k} \left|\widehat{f}(S)\right|$ refer to the level k Fourier mass of f.

Given functions $f, g : \{-1,1\}^n \to \mathbb{R}$, their convolution $f * g : \{-1,1\}^n \to \mathbb{R}$ is defined as $f * g(x) := \mathop{\mathbb{E}}_{y \sim U_n}[f(y)g(y \cdot x)]$. A standard fact about convolution of functions is that $\widehat{f * g}(S) = \widehat{f}(S)\widehat{g}(S)$ for all $S \subseteq [n]$.

2.2 Quantum Computation

Let $\mathcal{H}_m$ be the Hilbert space of dimension 2^m defined by the complex span of the orthonormal basis $\{|x\rangle : x \in \{-1,1\}^m\}$. We sometimes express these basis elements by integers $\{|i\rangle : i \in [2^m]\}$ by the same correspondence as before.

Fix any universal set of gates for quantum computation. A quantum circuit $Q : \{-1,1\}^n \to \{-1,1\}^m$ of space S consists of a set of S registers, the first n of which are initialized to $|x\rangle$, the input, while the rest are initialized to $|1\rangle$. It further consists of a sequence of operators chosen from the universal set of gates, along with a description of which register they act on. The size of a circuit is the number of operators. The output of a circuit is defined to be the contents of the first m registers. Since we want the output to be Boolean, we assume that the circuit measures these registers and returns the outcome. Thus, a quantum circuit is inherently probabilistic.

We now describe quantum circuits with *query or oracle* access. In this model, all registers are initialized to $|1\rangle$ and the input $x \in \{-1,1\}^n$ is not written into the registers. Instead, it is compactly presented to the algorithm using a blackbox, a device which for every index $i \in [n]$, returns $x(i)|i\rangle$ when it is given $|i\rangle$ as input. More precisely, for every possible input $x \in \{-1,1\}^n$, the oracle to x is the linear operator $O_x : \mathcal{H}_{\lceil \log n \rceil} \to \mathcal{H}_{\lceil \log n \rceil}$ which maps the basis states $|i\rangle$ to $x_i|i\rangle$ whenever $i \in [n]$ and otherwise leaves it fixed. This indeed restricts to a unitary operation on pure states, as its action on the basis states is described by a diagonal $\{-1,1\}$-matrix. This serves as the quantum analogue of a classical oracle, which is a blackbox that returns $x(i)$ on input $i \in [n]$. A *quantum circuit with oracle access to inputs* is a quantum circuit that is allowed to use the O_x operator in addition to the usual operators, where x is the input to the computation. The *size* of the circuit is the total number of gates used from the universal gate set plus the number of oracle queries used. We say that an algorithm is *efficient*, if it is described by a circuit of size at most $\mathit{poly} \log n$ with oracle access to inputs. Note that it is possible to use the oracle O_x to explicitly write down the input x into n registers, however, this requires n oracle calls and n registers. It is often the case that this step is unnecessary.

2.3 Classical & Quantum Communication Complexity

Let $f : \{-1,1\}^n \times \{-1,1\}^m \to \{-1,1\}$ be a partial Boolean function. Alice (respectively Bob) receives a private input $x \in \{-1,1\}^n$ (respectively $y \in \{-1,1\}^m$) and the players' goal is to compute $f(x,y)$ if (x,y) is in the support of f, while exchanging as few bits as possible. An input (x,y) is said to be a YES (respectively NO) instance if $f(x,y) = -1$ (respectively if $f(x,y) = 1$). We assume familiarity with *bounded-error randomized* and *quantum communication complexity*. In quantum communication with *entanglement*, Alice and Bob are given m independent copies of the Bell state for some $m \in \mathbb{N}$. In this case, we say that Alice and Bob share m bits of entanglement. In the *simultaneous* model of communication, Alice and Bob are not allowed to exchange messages with each other. Instead, they are allowed one round of communication with a referee Charlie, to whom they can only send qubits. The referee then performs some quantum operation on the qubits he receives and returns a bit as the output. As before, a bounded-error simultaneous protocol computes f if for all (x,y) in the support of f, with probability at least $2/3$, the referee's output agrees with $f(x,y)$. The cost is the total number of qubits that Alice and Bob send the referee.

Note that in each of the above models of communication, every function $f : \{-1,1\}^n \times \{-1,1\}^m \to \{-1,1\}$ has communication cost at most $n+m$, since the players may simply reveal their entire inputs. Hence, a small cost protocol is one in which the communication cost is at most $poly\log(n+m)$.

A communication protocol is said to be *efficient* if it can be implemented by a small size circuit with oracle access O_x, O_y to the inputs x, y. Protocols with small communication cost are not necessarily efficient, as they may require computationally intensive processing on the messages, or they may require the players to make several probes into their inputs.

2.4 The Forrelation Distribution $\mathcal{D}$

Let $x \sim \mathcal{D}$ refer to a random variable x distributed according to the probability distribution $\mathcal{D}$. We use $\mathbb{P}_{\mathcal{D}}$ to refer to the probability measure associated with $\mathcal{D}$ and $\mathbb{P}_{x\sim\mathcal{D}}(E(x))$ to refer to the probability of event $E(x)$ when $x \sim \mathcal{D}$. For an event $E(x)$, we will denote by $\mathcal{D}|E(x)$ (respectively $\mathcal{D}|\neg E(x)$), the distribution $\mathcal{D}$ conditioned on the event $E(x)$ occurring (respectively, the event $E(x)$ not occurring). Let $\epsilon \geq 0$ be a parameter, $f(x) : \mathbb{R}^n \to \mathbb{R}$ a function and $\mathcal{D}$ a distribution on $\mathbb{R}^n$. We say that $\mathcal{D}$ fools f with error ϵ if $\left| \mathop{\mathbb{E}}_{x\sim U_n}[f(x)] - \mathop{\mathbb{E}}_{x\sim\mathcal{D}}[f(x)] \right| \leq \epsilon$.

Let $\mathcal{N}(\mu, \sigma^2)$ denote a Gaussian distribution of mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}_{\geq 0}$. We will repeatedly use the following standard facts about Gaussians.

- Gaussian Concentration inequality: For $X \sim \mathcal{N}(\mu, \sigma^2)$, we have $\mathbb{P}[|X-\mu| \geq a] \leq e^{-\frac{a^2}{2\sigma^2}}$.
- The sum $\sum_i X_i$ of independent Gaussians $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ is distributed according to $\mathcal{N}(\sum_i \mu_i, \sum_i \sigma_i^2)$.

Let $N = 2^n$. The Hadamard matrix H_N is an $N \times N$ unitary matrix. We let x and y in $\{0,1\}^n$ index rows and columns of H_N respectively. The entries of H_N are as follows.

$$H_N(x,y) := \begin{cases} \frac{1}{\sqrt{N}} & \text{if } \sum_i x_i y_i \mod 2 = 0 \\ \frac{-1}{\sqrt{N}} & \text{otherwise} \end{cases}$$

Let $x = (x_1, x_2)$ for $x_1, x_2 \in \{-1,1\}^N$. We define the forrelation of x as the correlation between the second half x_2 and the Hadamard transform of the first half x_1.

$$forr(x) := \left\langle \frac{1}{\sqrt{N}} H_N(x_1) \middle| \frac{1}{\sqrt{N}} x_2 \right\rangle$$

We state the definition of the forrelation distribution, as defined in [15]. Fix a parameter $\epsilon = \frac{1}{50 \ln N}$. We first define an auxilliary Gaussian distribution $\mathcal{G}$ generated by sampling the first half uniformly at random and letting the second half be the Hadamard transform of the first half. More precisely,

1. Sample $x_1, \ldots, x_N \sim \mathcal{N}(0, \epsilon)$.
2. Let $y = H_N x$.
3. Output (x, y).

This is a Gaussian random variable in $2N$ dimensions of mean 0 and covariance matrix given by

$$\epsilon \begin{bmatrix} \mathbb{I}_N & H_N \\ H_N & \mathbb{I}_N \end{bmatrix}.$$

Let $trnc : \mathbb{R} \to [-1,1]$ be the truncation function which on input $\alpha > 1$, returns 1, $\alpha < -1$ returns -1 and otherwise returns α. This naturally defines a function $trnc : \mathbb{R}^{2N} \to [-1,1]^{2N}$ obtained by truncating each coordinate. We now define a distribution $\mathcal{D}$ over $\{-1,1\}^{2N}$ generated from $\mathcal{G}$ by truncating the sample and then independently sampling each coordinate as follows.

1. Sample $z \in \mathcal{G}$.
2. For each coordinate $i \in [2N]$ independently, let $z'_i = 1$ with probability $\frac{1+trnc(z_i)}{2}$ and -1 with probability $\frac{1-trnc(z_i)}{2}$.
3. Output z'.

We refer to the distribution $\mathcal{D}$ as the *forrelation* distribution. We state Claim 6.3 from [15] which implies that a vector drawn from this distribution has large forrelation on expectation. The proof is omitted.

▶ **Lemma 2.** *Let $\mathcal{D}$ be the forrelation distribution as defined previously. Then,*

$$\mathbb{E}_{z\sim\mathcal{D}}[forr(z)] \geq \frac{\epsilon}{2}.$$

2.5 Multilinear Functions on $\mathcal{D}$

Given a function $f : \{-1,1\}^n \to \mathbb{R}$, there is a unique multilinear polynomial $\tilde{f} : \mathbb{R}^n \to \mathbb{R}$ which agrees with f on $\{-1,1\}^n$. This polynomial is called the multilinear extension of f. The multilinear extension of any character function $\chi_S(x)$ is precisely $\prod_{i\in S} x_i$. The multilinear extension $\tilde{f}$ of f satisfies $\tilde{f}(x) = \sum_{S\subseteq[n]} \widehat{f}(S) \prod_{i\in S} x_i$ for all $x \in \mathbb{R}^n$. We sometimes identify f with its multilinear extension. The main content of this section is that bounded multilinear functions have similar expectations under $\mathcal{G}$ and under $\mathcal{D}$.

▷ Claim 3. Let $F : \mathbb{R}^{2N} \to \mathbb{R}$ be any multinear function $F = \sum_S \widehat{F}(S)\chi_S$. Then,

$$\mathop{\mathbb{E}}_{z'\sim\mathcal{D}}[F(z')] = \mathop{\mathbb{E}}_{z\sim\mathcal{G}}[F(trnc(z))].$$

The proof of this claim is identical to that of Equation (2) in [15]. The details can be found in the full version of this paper. The following claim states that $\mathbb{E}_{z\sim\mathcal{G}}[F(trnc(z))]$ is pretty close to $\mathbb{E}_{z\sim\mathcal{G}}[F(z)]$ for a bounded multilinear function F. Its proof is identical to that in [15], so we omit it. The underlying idea is that ϵ is small, so the random variable $z \sim \mathcal{G}$ has an exponentially decaying norm, furthermore, bounded multilinear functions F on $\{-1,1\}^{2N}$ cannot grow faster than exponentially in the norm of the argument.

▷ Claim 4. Let $F(z)$ be any multilinear polynomial mapping $\{-1,1\}^{2N}$ to $[-1,1]$. Let $z_0 \in [-1/2,1/2]^{2N}$, $p \leq \frac{1}{2}$ and $N > 1$. Then,

$$\mathbb{E}_{z\sim\mathcal{G}}\left[|F(trnc(z_0+pz)) - F(z_0+pz)|\right] \leq \frac{8}{N^5}.$$

We remark that the bound in [15] is $\frac{8}{N^2}$. The improved bound of $\frac{8}{N^5}$ in Claim 4 follows from our choice of $\epsilon = \frac{1}{50\ln N}$, as opposed to $\epsilon = \frac{1}{24\ln N}$ as in [15].

2.6 Moments of $\mathcal{G}$

In this section we state some facts about the moments of the forrelation distribution that will be useful later. We use the following notation to refer to the moments of $\mathcal{G}$.

$$\widehat{\mathcal{G}}(S,T) := \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{G}}\left[\prod_{i\in S} x_i \prod_{j\in T} y_j\right]$$

The following claim and its proof are analogous to Claim 4.1 in [15].

▷ Claim 5. Let $S, T \subseteq [N]$ and $i, j \in [N]$. Let $k_1 = |S|, k_2 = |T|$. Then,

1. $\widehat{\mathcal{G}}(\{i\}, \{j\}) = \epsilon N^{-1/2}(-1)^{\langle i,j \rangle}$.
2. $\widehat{\mathcal{G}}(S, T) = 0$ if $k_1 \neq k_2$.
3. $\left|\widehat{\mathcal{G}}(S, T)\right| \leq \epsilon^k k! N^{-k/2}$ if $k = k_1 = k_2$.
4. $\left|\widehat{\mathcal{G}}(S, T)\right| \leq \epsilon^{|S|}$ for all S, T.

3 The Forrelation Communication Problem

In this section we formally state the main theorems of this paper.

Let $\epsilon = \frac{1}{50 \ln N}$ be the parameter as before, defining the forrelation distribution. We restate Theorem 3.

▶ **Theorem 3.** *Consider the following distribution. A string $z \in \{-1, 1\}^{2N}$ is drawn from the forrelation distribution, $x \sim U_{2N}$ is drawn uniformly and $y := x \cdot z$. Alice gets x and Bob gets y. Given any deterministic communication protocol $C : \{-1, 1\}^{2N} \times \{-1, 1\}^{2N} \to \{-1, 1\}$ of cost $c \geq 1$, its expectation when the inputs are drawn from this distribution is close to when the inputs are drawn from the uniform distribution. That is,*

$$\left| \mathop{\mathbb{E}}_{\substack{x \sim U_{2N} \\ z \sim \mathcal{D}}} [C(x, x \cdot z)] - \mathop{\mathbb{E}}_{x,y \sim U_{2N}} [C(x, y)] \right| \leq O\left(\frac{c^2}{N^{1/2}}\right).$$

In other words, no deterministic protocol of cost $o(N^{1/4})$ has considerable advantage in distinguishing the above distribution from the uniform distribution.

▶ **Definition 6** (The Forrelation Problem)**.** *Alice is given $x \in \{-1, 1\}^{2N}$ and Bob is given $y \in \{-1, 1\}^{2N}$. Their goal is to compute the partial boolean function F defined as follows.*

$$F(x, y) = \begin{cases} -1 & \text{if } forr(x \cdot y) \geq \epsilon/4 \\ 1 & \text{if } forr(x \cdot y) \leq \epsilon/8. \end{cases}$$

We restate Theorem 1 and Theorem 2.

▶ **Theorem 1.** *The forrelation problem can be solved in the quantum simultaneous with entanglement model with $O(\log^3 N)$ bits of communication, when Alice and Bob are given access to $O(\log^3 N)$ bits of shared entanglement. Moreover, the protocol is efficient, as it can be implemented by a $O(\log^3 N)$ size quantum circuit with oracle access to inputs.*

The upper bound on the quantum communication complexity of the forrelation problem follows from the fact that the XOR of the inputs can be computed by a simultaneous-message quantum protocol, when the players share entanglement, and the fact that $forr(\circ)$ can be estimated by a small size quantum circuit [1, 2, 15]. The proof of Theorem 1 can be found in the full version of this paper.

▶ **Theorem 2.** *The randomized bounded-error interactive communication cost of the forrelation problem is $\tilde{\Omega}(N^{\frac{1}{4}})$.*

The lower bound on the randomized communication complexity of the forrelation problem follows from Theorem 3 and Lemma 2. The proof of Theorem 2 can be found in the full version of this paper. We now describe the proof of Theorem 3.

4 Proof of Theorem 3 : Distributional Lower Bound

Let $C : \{-1,1\}^{2N} \times \{-1,1\}^{2N} \to \{-1,1\}$ be any deterministic protocol of cost at most c. Let $D : \{-1,1\}^{2N} \times \{-1,1\}^{2N} \to \{-1,1\}$ be defined as follows. For $x, z \in \{-1,1\}^{2N}$,

$$D(x,z) := C(x, x \cdot z).$$

We will also use $D(x,z)$ to refer to its mulilinear extension. Note that our goal is to show that the function $\mathop{\mathbb{E}}_{x \sim U_{2N}}[D(x,z)]$ of z is fooled by $\mathcal{D}$. Towards this, we will prove that it is fooled by $p\mathcal{G}$ for small p. This approach was first used in [6] and is analogous to Claim 7.2 in [15].

▶ **Lemma 7.** *Let $p \leq \frac{1}{2N}$ and let $C(x,y)$ be any deterministic protocol of cost $c \geq 1$ for the forrelation problem. As before, let $D(x,z) : \mathbb{R}^{2N} \times \mathbb{R}^{2N} \to \mathbb{R}$ refer to the multilinear extension of $C(x, x \cdot z)$. Let $P \in [-p,p]^{2N}$. Then,*

$$\left| \mathop{\mathbb{E}}_{\substack{z \sim P \cdot \mathcal{G} \\ x \sim U_{2N}}} [D(x,z)] - \mathop{\mathbb{E}}_{z,x \sim U_{2N}} [D(x,z)] \right| \leq \frac{120 \epsilon c^2 p^2}{\sqrt{N}} + p^4 N^3.$$

Proof of Lemma 7. We begin by observing some properties of the distribution $P \cdot \mathcal{G}$. The sample $z \sim P \cdot \mathcal{G}$ is obtained by scaling the i-th coordinate of $z' \sim \mathcal{G}$ by P_i for each $i \in [2N]$. This implies that for all $S \subseteq [2N]$,

$$\mathop{\mathbb{E}}_{z \sim P \cdot \mathcal{G}} [\chi_S(z)] = \left(\prod_{i \in S} P_i \right) \mathop{\mathbb{E}}_{z \sim \mathcal{G}} [\chi_S(z)] \,. \tag{1}$$

Part (2.) of Claim 5 implies that the odd moments of $\mathcal{G}$ are zero. Equation (1) implies that this is also true for $P \cdot \mathcal{G}$. That is, for all $S \subseteq [2N]$,

$$|S| \text{ is odd} \implies \mathop{\mathbb{E}}_{z \sim P \cdot \mathcal{G}} [\chi_S(z)] = 0. \tag{2}$$

Part (3.) of Claim 5 implies that for $S \subseteq [2N], |S| = 2k$, the S-th moment $\mathbb{E}_{z \sim \mathcal{G}} \chi_S(z)$ is at most $\epsilon^k k! N^{-k/2}$ in magnitude. Along with equation (1), this implies that for $k \in \mathbb{N}$,

$$|S| = 2k \implies \left| \mathop{\mathbb{E}}_{z \sim P \cdot \mathcal{G}} [\chi_S(z)] \right| \leq \left(\prod_{i \in S} P_i \right) \epsilon^k k! N^{-k/2} \leq p^{2k} \epsilon^k k! N^{-k/2}. \tag{3}$$

We now proceed with the proof of the lemma. Let

$$\Delta := \left| \mathop{\mathbb{E}}_{\substack{z \sim P \cdot \mathcal{G} \\ x \sim U_{2N}}} [D(x,z)] - \mathop{\mathbb{E}}_{z,x \sim U_{2N}} [D(x,z)] \right|.$$

Note that this is the quantity we wish to bound in the lemma. For ease of notation, let $H : \{-1,1\}^{2N} \to [-1,1]$ be defined at every point $z \in \{-1,1\}^{2N}$ by

$$H(z) := \mathop{\mathbb{E}}_{x \sim U_{2N}} [D(x,z)] \,.$$

We identify $H(z)$ with its multilinear extension. Note that by uniqueness of multilinear extensions, the above equality holds even for $z \in \mathbb{R}^{2N}$. This implies that

$$\mathop{\mathbb{E}}_{\substack{z \sim P \cdot \mathcal{G} \\ x \sim U_{2N}}} [D(x,z)] = \mathop{\mathbb{E}}_{z \sim P \cdot \mathcal{G}} [H(z)] \quad \text{and} \quad \mathop{\mathbb{E}}_{z,x \sim U_{2N}} [D(x,z)] = \mathop{\mathbb{E}}_{z \sim U_{2N}} [H(z)].$$

This, along with the definition of Δ implies that

$$\Delta = \left| \mathop{\mathbb{E}}_{z \sim P \cdot \mathcal{G}}[H(z)] - \mathop{\mathbb{E}}_{z \sim U_{2N}}[H(z)] \right|.$$

Note that $H(z) = \sum_S \widehat{H}(S)\chi_S(z)$ for all $z \in \mathbb{R}^{2N}$. This implies that for all distributions $\mathcal{Z}$ on $\mathbb{R}^{2N}$, we have $\mathop{\mathbb{E}}_{z \sim \mathcal{Z}}[H(z)] = \sum_S \widehat{H}(S) \mathop{\mathbb{E}}_{z \sim \mathcal{Z}}[\chi_S(z)]$. This implies that

$$\Delta = \left| \sum_{S \subseteq [2N]} \widehat{H}(S) \left(\mathop{\mathbb{E}}_{z \sim P \cdot \mathcal{G}}[\chi_S(z)] - \mathop{\mathbb{E}}_{z \sim U_{2N}}[\chi_S(z)] \right) \right|.$$

For any probability distribution, the moment corresponding to the empty set is 1 by definition. For all non empty sets S, we have $\mathop{\mathbb{E}}_{z \sim U_{2N}}[\chi_S(z)] = 0$. Using this fact in the above equality, along with the triangle inequality, we have

$$\Delta = \left| \sum_{\emptyset \neq S \subseteq [2N]} \widehat{H}(S) \mathop{\mathbb{E}}_{z \sim P \cdot \mathcal{G}}[\chi_S(z)] \right| \leq \sum_{\emptyset \neq S \subseteq [2N]} \left| \widehat{H}(S) \right| \left| \mathop{\mathbb{E}}_{z \sim P \cdot \mathcal{G}}[\chi_S(z)] \right|.$$

We use the bounds from (2) and (3) on the moments of $P \cdot \mathcal{G}$ to derive the following.

$$\begin{aligned} \Delta &\leq \sum_{\substack{|S|=2k \\ k \geq 1}} \left| \widehat{H}(S) \right| p^{2k} \epsilon^k k! N^{-k/2} \\ &= \sum_{k \geq 1} L_{2k}(H) p^{2k} \epsilon^k k! N^{-k/2} \end{aligned}$$

We upper bound $L_{2k}(H)$ by $\binom{2N}{2k}$ when $k \geq 2$. This implies that

$$\begin{aligned} \Delta &\leq L_2(H) \frac{\epsilon p^2}{\sqrt{N}} + \sum_{k \geq 2} \binom{2N}{2k} p^{2k} \epsilon^k k! N^{-k/2} \\ &\leq L_2(H) \frac{\epsilon p^2}{\sqrt{N}} + \sum_{k \geq 2} \frac{2^{2k} N^{2k}}{(2k)!} p^{2k} \epsilon^k k! N^{-k/2} \\ &\leq L_2(H) \frac{\epsilon p^2}{\sqrt{N}} + \sum_{k \geq 2} N^{3k/2} p^{2k} 4^k \epsilon^k. \end{aligned}$$

In the summation $\sum_{k \geq 2} N^{3k/2} p^{2k} 4^k \epsilon^k$, we see that every successive term is smaller than the previous by a factor of at least $1/4$. This is because the assumption $p \leq \frac{1}{2N}$ implies that $p^2 N^{3/2} \leq p^2 N^2 \leq \frac{1}{4}$ and because $4\epsilon \leq 1$. Thus, we can bound this summation by twice the first term, which is $16 p^4 N^3 \epsilon^2$. This implies that

$$\Delta \leq L_2(H) \frac{\epsilon p^2}{\sqrt{N}} + 32 p^4 N^3 \epsilon^2.$$

Since $\epsilon = \frac{1}{50 \ln N} \leq \frac{1}{32}$, we may bound $32 p^4 N^3 \epsilon^2$ by $p^4 N^3$. This implies that

$$\Delta \leq L_2(H) \frac{\epsilon p^2}{\sqrt{N}} + p^4 N^3.$$

We restate Claim 1 which provides a bound on $L_2(H)$.

▷ Claim 1. Let $C(x,y) : \{-1,1\}^{2N} \times \{-1,1\}^{2N} \to \{-1,1\}$ be any deterministic protocol of cost $c \geq 1$, let $D(x,z) : \mathbb{R}^{2N} \times \mathbb{R}^{2N} \to \mathbb{R}$ refer to the unique multilinear extension of $C(x, x \cdot z)$ and $H : \mathbb{R}^{2N} \to \mathbb{R}$ be defined by $H(z) = \mathbb{E}_{x \sim U_{2N}} D(x,z)$. Then,

$$L_2(H) \triangleq \sum_{|S|=2} |\widehat{H}(S)| \leq 120c^2.$$

This claim along with the preceding inequality implies that

$$\Delta \leq 120c^2 \frac{\epsilon p^2}{\sqrt{N}} + p^4 N^3.$$

This completes the proof of Lemma 7. ◀

Proof of Claim 1. In order to bound the level-2 Fourier mass of H, we will use the following lemma. Its statement and proof appear as "Level-k Inequalities" on Page 259 of "Analysis of Boolean Functions" [12].

▶ **Lemma 8** (Level-k Inequalities). *Let $F : \{-1,1\}^n \to \{0,1\}$ have mean $\mathbb{E}[F] = \alpha$ and let $k \in \mathbb{N}$ be at most $2\ln(1/\alpha)$. Then,*

$$\sum_{|S|=k} \left(\widehat{F}(S)\right)^2 \leq \alpha^2 \left(\frac{2e}{k} \ln(1/\alpha)\right)^k.$$

We now show the desired bound on $L_2(H)$. Since C is a deterministic protocol of cost at most c, it induces a partition of the input space $\{-1,1\}^{2N} \times \{-1,1\}^{2N}$ into at most 2^c rectangles. Let $\mathcal{P}$ be this partition and let $A \times B$ index rectangles in $\mathcal{P}$, where A (respectively B) is the set of Alice's (respectively Bob's) inputs compatible with the rectangle. Let $C(A \times B) \in \{-1,1\}$ be the output of the protocol on inputs from a rectangle $A \times B \in \mathcal{P}$. For all $x, y \in \{-1,1\}^{2N}$, we have

$$C(x,y) = \sum_{A \times B \in \mathcal{P}} C(A \times B) \mathbb{1}_A(x) \mathbb{1}_B(y).$$

By definition, $D(x,z) = C(x, x \cdot z)$. This implies that

$$D(x,z) = \sum_{A \times B \in \mathcal{P}} C(A \times B) \mathbb{1}_A(x) \mathbb{1}_B(x \cdot z).$$

Taking an expectation over $x \sim U_{2N}$ of the above identity implies that

$$H(z) \triangleq \mathop{\mathbb{E}}_{x \sim U_{2N}}[D(x,z)] = \sum_{A \times B \in \mathcal{P}} C(A \times B)\big(\mathbb{1}_A * \mathbb{1}_B\big)(z).$$

This implies that for any $S \subseteq [n]$, we have

$$\widehat{H}(S) = \sum_{A \times B \in \mathcal{P}} C(A \times B) \widehat{\mathbb{1}_A * \mathbb{1}_B}(S) = \sum_{A \times B \in \mathcal{P}} C(A \times B) \widehat{\mathbb{1}_A}(S) \widehat{\mathbb{1}_B}(S).$$

We thus obtain

$$\begin{aligned} L_2(H) &= \sum_{|S|=2} \left|\widehat{H}(S)\right| \\ &= \sum_{|S|=2} \left| \sum_{A \times B \in \mathcal{P}} C(A \times B) \widehat{\mathbb{1}_A}(S) \widehat{\mathbb{1}_B}(S) \right| \\ &\leq \sum_{A \times B \in \mathcal{P}} \sum_{|S|=2} |\widehat{\mathbb{1}_A}(S)| |\widehat{\mathbb{1}_B}(S)|. \end{aligned}$$

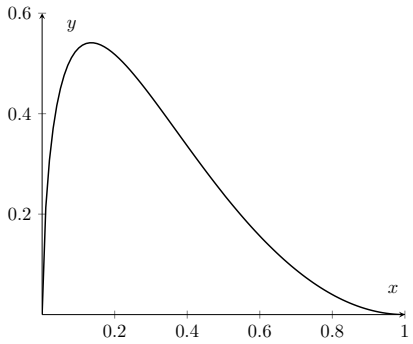


Figure 1 Plot of the function $y = x\left(\ln \frac{1}{x}\right)^2$.

We apply Cauchy Schwarz to the term $\sum_{|S|=2} |\widehat{\mathbb{1}_A}(S)||\widehat{\mathbb{1}_B}(S)|$ to obtain

$$L_2(H) \leq \sum_{A\times B\in\mathcal{P}} \Big(\sum_{|S|=2} \widehat{\mathbb{1}_A}(S)^2 \Big)^{1/2} \Big(\sum_{|S|=2} \widehat{\mathbb{1}_B}(S)^2 \Big)^{1/2}.$$

For ease of notation, let $\mu(A) = \frac{|A|}{2^{2N}}$ denote the measure of a set $A \subseteq \{-1,1\}^{2N}$ under U_{2N}. We first ensure that for each rectangle $A \times B \in \mathcal{P}$, we have $\mu(A) \leq \frac{1}{e}$ and $\mu(B) \leq \frac{1}{e}$. We may do this by adding 2 extra bits of communication for each player. For $k = 2$, we have $k = 2\ln(e) \leq 2\ln \frac{1}{\mu(A)}$ and $k \leq 2\ln \frac{1}{\mu(B)}$. We apply Lemma 8 on the indicator functions $\mathbb{1}_A$ and $\mathbb{1}_B$ for $k = 2$ to obtain

$$\sum_{|S|=2} \Big(\widehat{\mathbb{1}_A}(S)\Big)^2 \leq \mu(A)^2 \Big(e\ln(1/\mu(A))\Big)^2 \quad \text{and} \quad \sum_{|S|=2} \Big(\widehat{\mathbb{1}_B}(S)\Big)^2 \leq \mu(B)^2 \Big(e\ln(1/\mu(B))\Big)^2.$$

Substituting this in the bound for $L_2(H)$, we have

$$L_2(H) \leq e^2 \sum_{A\times B\in\mathcal{P}} \mu(A)\mu(B) \ln\frac{1}{\mu(A)} \ln\frac{1}{\mu(B)}.$$

Let $\Delta := e^2 \sum_{A\times B\in\mathcal{P}} \mu(A)\mu(B) \ln \frac{1}{\mu(A)} \ln \frac{1}{\mu(B)}$ be the expression in the R.H.S. of the above. Note that it suffices to upper bound Δ. Consider the case when $\mathcal{P}$ consists of 2^c rectangles $A \times B$, each of which satisfies $\mu(A) = \mu(B) = \frac{1}{2^{c/2}}$. In this case, Δ evaluates to $e^2 \sum_{A\times B\in\mathcal{P}} \frac{1}{2^c} (\frac{c\ln 2}{2})^2 = O(c^2)$. This proves the lemma in this special case. A similar bound holds for the general case and the proof follows from a concavity argument that we describe now.

Since $\mu(A), \mu(B) \leq 1$, we have the following inequality.

$$\begin{aligned}
\Delta &\triangleq e^2 \sum_{A\times B\in\mathcal{P}} \mu(A)\mu(B) \ln\frac{1}{\mu(A)} \ln\frac{1}{\mu(B)} \\
&\leq e^2 \sum_{A\times B\in\mathcal{P}} \mu(A)\mu(B) \ln\frac{1}{\mu(A)\mu(B)} \ln\frac{1}{\mu(A)\mu(B)} \\
&= e^2 \sum_{A\times B\in\mathcal{P}} \mu(A\times B) \left(\ln\frac{1}{\mu(A\times B)}\right)^2.
\end{aligned}$$

Let $f : [0, \infty) \to \mathbb{R}$ be defined by $f(p) := p \ln(1/p)^2$. A small calculation shows that f is a concave function in the interval $[0, 0.3]$ (see Figure 7). Let $\alpha_i \in [0, 0.3]$ for $i \in [k]$. Jensen's inequality applied to f states that for $i \sim [k]$ drawn uniformly at random, we have $\mathbb{E}_i[f(\alpha_i)] \leq f(\mathbb{E}_i[\alpha_i])$. This implies that

$$\sum_{i=1}^{k} \alpha_i \ln(1/\alpha_i)^2 \leq \left(\sum_{i=1}^{k} \alpha_i\right) \ln\left(\frac{k}{\sum_{i=1}^{k} \alpha_i}\right)^2.$$

We apply this inequality to the terms in Δ by substituting α_i with $\mu(A \times B)$. We may do this since the assumption that $\mu(A), \mu(B) \leq \frac{1}{e}$ implies that $\mu(A \times B) \leq \frac{1}{e^2} \leq 0.3$. This implies that

$$\Delta \leq e^2 \left(\sum_{A \times B \in \mathcal{P}} \mu(A \times B)\right) \ln\left(\frac{2^{c+4}}{\sum_{A \times B \in \mathcal{P}} \mu(A \times B)}\right)^2$$

Since $\sum_{A \times B \in \mathcal{P}} \mu(A \times B) = 1$, we have

$$\Delta \leq e^2 (c+4)^2 (\ln 2)^2 \leq 120 c^2.$$

This completes the proof of Claim 1. ◁

We now show that an analogue of Lemma 7 holds for restricted protocols, similarly to Claim 7.3 in [15].

▶ **Lemma 9.** *Let $p \leq \frac{1}{4N}$ and $C(x, y)$ be any deterministic protocol of cost $c \geq 1$ for the forrelation problem. As before, let $D(x, z) : \mathbb{R}^{2N} \times \mathbb{R}^{2N} \to \mathbb{R}$ refer to the multilinear extension of $C(x, x \cdot z)$. Let $z_0 \in [-1/2, 1/2]^{2N}$. Then,*

$$\left| \mathop{\mathbb{E}}_{\substack{z \sim p\mathcal{G} \\ x \sim U_{2N}}} [D(x, z_0 + z)] - \mathop{\mathbb{E}}_{z, x \sim U_{2N}} [D(x, z_0 + z)] \right| \leq \frac{120 \epsilon c^2 (2p)^2}{\sqrt{N}} + (2p)^4 N^3.$$

▶ **Corollary 10.** *Under the same hypothesis as in Lemma 9,*

$$\left| \mathbb{E}_{z \sim p\mathcal{G}} [D(0, z_0 + z)] - D(0, z_0) \right| \leq \frac{120 \epsilon c^2 (2p)^2}{\sqrt{N}} + (2p)^4 N^3.$$

Proof of Corollary 10 from Lemma 9. Since $D(x, z)$ is a multilinear polynomial, for all $z \in \mathbb{R}^{2N}$, we have $\mathbb{E}_{x \sim U_{2N}}[D(x, z)] = D(0, z)$. This implies that for all $z_0 \in \mathbb{R}^{2N}$,

$$\mathop{\mathbb{E}}_{\substack{z \sim p\mathcal{G} \\ x \sim U_{2N}}} [D(x, z_0 + z)] = \mathbb{E}_{z \sim p\mathcal{G}} [D(0, z_0 + z)].$$

For all $z_0 \in \mathbb{R}^{2N}$, since $\mathbb{E}_{z \sim U_{2N}}[D(0, z_0 + z)] = D(0, z_0)$, we have

$$\mathop{\mathbb{E}}_{z, x \sim U_{2N}} [D(x, z_0 + z)] = D(0, z_0).$$

The proof of Corollary 10 follows from the above two equalities and Lemma 9. ◀

Proof of Lemma 9. Similarly to the approach of [6, 15], we will express $D(x, z_0 + z)$ as the average output of restricted protocols $(C \circ \rho)(x, x \cdot z)$, on which we can use Lemma 7 to derive the result. These restricted protocols roughly correspond to Alice and Bob fixing a common subset $I \subseteq [2N]$ of their inputs in a predetermined way and then running the original protocol. We formalize this now.

A restriction ρ of $\mathbb{R}^{2N}$ is an element of $\{-1,1,*\}^{2N}$. It defines an action $\rho : \mathbb{R}^{2N} \to \mathbb{R}^{2N}$ in the following natural way. For any $z \in \mathbb{R}^{2N}$ and $i \in [2N]$,

$$(\rho(z))(i) := \begin{cases} \rho(i) & \text{if } \rho(i) \in \{-1,1\} \\ z(i) & \text{otherwise.} \end{cases}$$

Let $sign : (\mathbb{R} \setminus 0) \to \{-1,1\}$ be the function which maps real numbers to their sign. Given $z_0 \in [-1/2,1/2]^{2N}$, let R_{z_0} be a distribution over restrictions of $\mathbb{R}^{2N}$ defined as follows. For each $i \in [2N]$, independently, set[1]:

$$\rho(i) := \begin{cases} sign(z_0(i)) & \text{with probability } |z_0(i)| \\ * & \text{with probability } 1 - |z_0(i)|. \end{cases}$$

Let $P \in \mathbb{R}^{2N}$ be such that $P_i := \frac{1}{1-|z_0(i)|}$ for every $i \in [2N]$. Note that the assumption of $z_0 \in [-1/2,1/2]^{2N}$ ensures that P is a well defined element of $[1,2]^{2N}$. For any $z \in \mathbb{R}^{2N}$ and $i \in [2N]$, the expected value of the ith coordinate of $\rho(z)$ when $\rho \sim R_{z_0}$ can be computed as follows.

$$\mathop{\mathbb{E}}_{\rho \sim R_{z_0}} [(\rho(z))(i)] = |z_0(i)| sign(z_0(i)) + (1 - |z_0(i)|) z(i) = z_0(i) + \frac{1}{P_i} z(i)$$

This implies that for any fixed $x, z \in \mathbb{R}^{2N}$ and $z_0 \in [-1/2,1/2]^{2N}$, since D is a multilinear function, we have

$$\mathop{\mathbb{E}}_{\rho \sim R_{z_0}} [D(x, \rho(z))] = D(x, \mathop{\mathbb{E}}_{\rho \sim R_{z_0}} [\rho(z)]) = D(x, z_0 + P^{-1} \cdot z).$$

Replacing z with $P \cdot z$ in the above equality implies that

$$\mathop{\mathbb{E}}_{\rho \sim R_{z_0}} [D(x, \rho(P \cdot z))] = D(x, z_0 + z).$$

This equality allows us to rewrite the L.H.S. of Lemma 9 as follows.

$$\begin{aligned} \Delta &:= \left| \mathop{\mathbb{E}}_{\substack{z \sim pG, \\ x \sim U_{2N}}} [D(x, z_0 + z)] - \mathop{\mathbb{E}}_{z,x \sim U_{2N}} [D(x, z_0 + z)] \right| \\ &= \left| \mathop{\mathbb{E}}_{\substack{z \sim pP \cdot \mathcal{G}, \rho \sim R_{z_0} \\ x \sim U_{2N}}} [D(x, \rho(z))] - \mathop{\mathbb{E}}_{\substack{z \sim P \cdot U_{2N}, \rho \sim R_{z_0} \\ x \sim U_{2N}}} [D(x, \rho(z))] \right| \\ &= \left| \mathop{\mathbb{E}}_{\rho \sim R_{z_0}} \left[\mathop{\mathbb{E}}_{\substack{z \sim pP \cdot \mathcal{G}, \\ x \sim U_{2N}}} [D(x, \rho(z))] - \mathop{\mathbb{E}}_{\substack{z \sim P \cdot U_{2N}, \\ x \sim U_{2N}}} [D(x, \rho(z))] \right] \right|. \end{aligned}$$

For a multilinear polynomial, its expectation over a product distribution depends only on the mean of that distribution. This allows us to replace the expectation of $D(x, \rho(z))$ over $z \sim P \cdot U_{2N}$ by an expectation over $z \sim U_{2N}$. We thus obtain

$$\Delta = \left| \mathop{\mathbb{E}}_{\rho \sim R_{z_0}} \left[\mathop{\mathbb{E}}_{\substack{z \sim pP \cdot \mathcal{G}, \\ x \sim U_{2N}}} [D(x, \rho(z))] - \mathop{\mathbb{E}}_{\substack{z \sim U_{2N}, \\ x \sim U_{2N}}} [D(x, \rho(z))] \right] \right| \tag{4}$$

[1] If $z_0(i)$ is zero, then $\rho(i) = *$ with probability 1.

For any $\rho \in \{-1,1,*\}^{2N}$ and $u \in \{-1,1\}^{2N}$, we define a substitution $\rho^u : \mathbb{R}^{2N} \to \mathbb{R}^{2N}$ obtained from ρ and u as follows. For any $x \in \mathbb{R}^{2N}$ and $i \in [2N]$,

$$(\rho^u(x))(i) := \begin{cases} u(i) & \text{if } \rho(i) \in \{-1,1\} \\ x(i) & \text{otherwise.} \end{cases}$$

This is an action on $\mathbb{R}^{2N}$ which replaces the values of coordinates specified by ρ, with values from u. For every fixed ρ, as we vary over $x, u \sim U_{2N}$ the distribution of $\rho^u(x)$ is exactly U_{2N}. This implies that for all $z \in \mathbb{R}^{2N}, \rho \in \{-1,1,*\}^{2N}$,

$$\mathop{\mathbb{E}}_{x\sim U_{2N}}[D(x,\rho(z))] = \mathop{\mathbb{E}}_{x,u\sim U_{2N}}[D(\rho^u(x),\rho(z))].$$

Substituting this in equation (4), we have

$$\Delta = \left| \mathop{\mathbb{E}}_{\rho\sim R_{z_0}} \mathop{\mathbb{E}}_{u\sim U_{2N}} \left[\mathop{\mathbb{E}}_{\substack{z\sim pP\cdot\mathcal{G},\\ x\sim U_{2N}}}[D(\rho^u(x),\rho(z))] - \mathop{\mathbb{E}}_{z,x\sim U_{2N}}[D(\rho^u(x),\rho(z))] \right] \right|.$$

Applying Triangle Inequality on the above, we have

$$\Delta \leq \mathop{\mathbb{E}}_{\rho\sim R_{z_0}} \mathop{\mathbb{E}}_{u\sim U_{2N}} \left| \mathop{\mathbb{E}}_{\substack{z\sim pP\cdot\mathcal{G},\\ x\sim U_{2N}}}[D(\rho^u(x),\rho(z))] - \mathop{\mathbb{E}}_{z,x\sim U_{2N}}[D(\rho^u(x),\rho(z))] \right|. \tag{5}$$

Fix any $\rho \in \{-1,1,*\}^{2N}$ and $u \in \{-1,1\}^{2N}$. For every $x, z \in \{-1,1\}^{2N}$, we have $D(x,z) = C(x, x\cdot z)$, furthermore, $\rho^u(x), \rho(z) \in \{-1,1\}^{2N}$. This implies that for every $x, z \in \{-1,1\}^{2N}$,

$$D(\rho^u(x),\rho(z)) = C(\rho^u(x), \rho^u(x)\cdot\rho(z)). \tag{6}$$

This prompts us to define a communication protocol $C \circ \rho^u$ where Alice and Bob first restrict their inputs and then run the original protocol C. The restriction is that for each coordinate $i \in [2N]$ with $\rho_i \in \{-1,1\}$, Alice overwrites her input x_i with u_i while Bob overwrites his input y_i with $\rho_i u_i$. The main property of this restricted protocol is that for all $x, z \in \{-1,1\}^{2N}$,

$$(C\circ\rho^u)(x, x\cdot z) = C(\rho^u(x), \rho^u(x)\cdot\rho(z)).$$

This, along with equation (6) implies that $D(\rho^u(x),\rho(z))$ is the unique multilinear extension of $(C\circ\rho^u)(x, x\cdot z)$. The cost of $C\circ\rho^u$ is at most that of C since Alice and Bob don't need to communicate to restrict their inputs. We now use Lemma 7 on $C\circ\rho^u$ to argue that $pP\cdot\mathcal{G}$ fools $\mathop{\mathbb{E}}_{x\sim U_{2N}}[D(\rho^u(x),\rho(z))]$. The conditions of the lemma are satisfied since $pP \in [-2p,2p]^{2N}$, $p \leq \frac{1}{4N}$, and $C\circ\rho^u$ is a protocol of cost at most c and whose multilinear extension is $D(\rho^u(x),\rho(z))$. The lemma implies that

$$\left| \mathop{\mathbb{E}}_{\substack{z\sim pP\cdot\mathcal{G},\\ x\sim U_{2N}}}[D(\rho^u(x),\rho(z))] - \mathop{\mathbb{E}}_{\substack{z\sim U_{2N},\\ x\sim U_{2N}}}[D(\rho^u(x),\rho(z))] \right| \leq \frac{120\epsilon c^2(2p)^2}{\sqrt{N}} + (2p)^4 N^3.$$

Substituting this in inequality (5) completes the proof of Lemma 9. ◀

Proof of Theorem 3. Since $D(x,z)$ is the multilinear extension of $C(x, x\cdot z)$ and since $\mathcal{D}$ and U_{2N} are distributions over $\{-1,1\}^{2N}$, we have

$$\mathbb{E}_{x\sim U_{2N}, z\sim\mathcal{D}}[C(x,x\cdot z)] = \mathbb{E}_{x\sim U_{2N}, z\sim\mathcal{D}}[D(x,z)] = \mathbb{E}_{z\sim\mathcal{D}}[D(0,z)].$$

When $x\sim U_{2N}$ and $y\sim U_{2N}$ are independently sampled, the distribution of $(x, x\cdot y)$ is U_{4N}. This implies that

$$\mathbb{E}_{x,y\sim U_{2N}}[C(x,y)] = \mathbb{E}_{x,y\sim U_{2N}}[D(x,x\cdot y)] = D(0,0).$$

The above two equations allow us to rewrite the quantity in the L.H.S. of Theorem 3 as follows.

$$\Delta := \left| \mathop{\mathbb{E}}_{\substack{x\sim U_{2N}\\ z\sim\mathcal{D}}}[C(x,x\cdot z)] - \mathop{\mathbb{E}}_{x,y\sim U_{2N}}[C(x,y)] \right| = \left| \mathbb{E}_{z\sim\mathcal{D}}[D(0,z)] - D(0,0) \right|$$

Claim 3 applied on the multilinear polynomial D implies that $\mathbb{E}_{z\sim\mathcal{D}}[D(0,z)] = \mathbb{E}_{z\sim\mathcal{G}}[D(0,trnc(z))]$. Substituting this in the above equality implies that

$$\Delta = \left| \mathbb{E}_{z\sim\mathcal{G}}[D(0,trnc(z))] - D(0,0) \right|.$$

Let $t = 16N^4, p = \frac{1}{\sqrt{t}} = \frac{1}{4N^2}$. Let $z^{(1)},\ldots,z^{(t)}\sim\mathcal{G}$ be independent samples and let Z refer to this collection of random variables. For $i\in[t]$, define $z^{\le(i)} := p(z^{(1)}+\ldots+z^{(i)})$. By convention, $z^{\le(0)} := 0$. Note that for $i\in[t]$, $z^{\le(i)}$ has a Gaussian distribution with mean 0 and covariance matrix as $p^2 i$ times that of $\mathcal{G}$. Thus, $z^{\le(t)}$ is sampled according to $\mathcal{G}$. Substituting this in the previous equality implies that

$$\Delta = \left| \mathbb{E}_Z[D(0,trnc(z^{\le t}))] - D(0,0) \right|.$$

To bound the above quantity, for each $0\le i\le t-1$, we show a bound on

$$\Delta_i := \left| \mathbb{E}_Z[D(0,trnc(z^{\le(i+1)}))] - \mathbb{E}_Z[D(0,trnc(z^{\le(i)}))] \right|.$$

Since $z^{\le(0)}=0$, the triangle inequality implies that $\Delta\le\sum_{i=0}^{t-1}\Delta_i$.

Fix any $i\in\{0,\ldots,t-1\}$. We now bound Δ_i. Let E_i be the event that $z^{\le(i)}\notin[-1/2,1/2]^{2N}$. We first observe that E_i is a low probability event. Since each $z^{\le(i)}(j)$ is distributed as $\mathcal{N}(0,p^2 i\epsilon)$, where $p^2 i\le 1$ and $\epsilon = 1/(50\ln N)$, we have

$$\mathbb{P}[z^{\le(i)}(j)\notin[-1/2,1/2]] \le \mathbb{P}[|\mathcal{N}(0,\epsilon)|\ge 1/2] \le \exp(-1/8\epsilon) \le \exp(-6\ln N) = \frac{1}{N^6}$$

Applying a Union bound over coordinates $j\in[2N]$, we have for each $0\le i\le t$,

$$\mathbb{P}[E_i] = \mathbb{P}[z^{\le(i)}\notin[-1/2,1/2]^{2N}] \le 2N\frac{1}{N^6} \le \frac{2}{N^5}. \tag{7}$$

When E_i does not occur, we have $trnc(z^{\le(i)}) = z^{\le(i)}\in[-1/2,1/2]^{2N}$. For every fixed value of $z^{\le(i)}$ in this range, we apply Corollary 10 with parameters $p=\frac{1}{4N^2}$, $z_0 = z^{\le(i)}$ and $z = z^{\le(i+1)} - z^{\le(i)} = pz^{(i+1)}$. Note that the conditions in the hypothesis are satisfied since $z_0\in[-1/2,1/2]^{2N}$, $p\le 1/(4N)$ and the random variable $pz^{(i+1)}$ is distributed as $p\mathcal{G}$. The corollary implies that for every $z^{\le(i)}\in[-1/2,1/2]^{2N}$,

$$\left| \mathbb{E}_Z\left[D(0,z^{\le(i+1)}) \mid z^{\le(i)} \right] - \mathbb{E}_Z\left[D(0,z^{\le(i)}) \mid z^{\le(i)} \right] \right| \le \frac{120\epsilon c^2(2p)^2}{N^{1/2}} + (2p)^4 N^3.$$

Since $\neg E_i$ implies that $z^{\leq(i)} \in [-1/2, 1/2]^{2N}$, we have

$$\left|\mathbb{E}_Z\left[D(0, z^{\leq(i+1)}) \mid \neg E_i\right] - \mathbb{E}_Z\left[D(0, z^{\leq(i)}) \mid \neg E_i\right]\right| \leq \frac{120\epsilon c^2 (2p)^2}{N^{1/2}} + (2p)^4 N^3.$$

We apply Claim 4 on the multilinear polynomial $D(0, z) : [-1, 1]^{2N} \to [-1, 1]$ with the parameters $p = \frac{1}{4N^2}$, $z_0 = z^{\leq(i)}$ and $z = z^{(i+1)}$. Note that the conditions are satisfied since $z_0 \in [1/2, 1/2]^{2N}$ and $p \leq \frac{1}{2}$. The claim implies that

$$\left|\mathbb{E}_Z\left[D(0, z^{\leq(i+1)}) \mid \neg E_i\right] - \mathbb{E}_Z\left[D(0, trnc(z^{\leq(i+1)})) \mid \neg E_i\right]\right| \leq \frac{8}{N^5}.$$

The previous two inequalities, along with the triangle inequality, imply that

$$\left|\mathbb{E}_Z\left[D(0, trnc(z^{\leq(i+1)})) \mid \neg E_i\right] - \mathbb{E}_Z\left[D(0, z^{\leq(i)}) \mid \neg E_i\right]\right| \leq \frac{120\epsilon c^2 (2p)^2}{N^{1/2}} + (2p)^4 N^3 + \frac{8}{N^5}. \tag{8}$$

Note that for every possible values of $z^{\leq(i+1)}$ and $z^{\leq(i)}$, the difference $D(0, trnc(z^{\leq(i+1)})) - D(0, trnc(z^{\leq(i)}))$ is bounded in magnitude by 2, since $D(0, trnc(z))$ maps $\mathbb{R}^{2N}$ to $[-1, 1]$. This implies that

$$\left|\mathbb{E}_Z\left[D(0, trnc(z^{\leq(i+1)})) \mid E_i\right] - \mathbb{E}_Z\left[D(0, trnc(z^{\leq(i)})) \mid E_i\right]\right| \leq 2.$$

Thus, we have

$$\begin{aligned}
\Delta_i &\leq \mathbb{P}[\neg E_i] \cdot \left|\mathbb{E}_Z[D(0, trnc(z^{\leq(i+1)})) \mid \neg E_i] - \mathbb{E}_Z[D(0, trnc(z^{\leq(i)})) \mid \neg E_i]\right| \\
&\quad + \mathbb{P}[E_i] \cdot \left|\mathbb{E}_Z[D(0, trnc(z^{\leq(i+1)})) \mid E_i] - \mathbb{E}_Z[D(0, trnc(z^{\leq(i)})) \mid E_i]\right| \\
&\leq \left|\mathbb{E}_Z[D(0, trnc(z^{\leq(i+1)})) \mid \neg E_i] - \mathbb{E}_Z[D(0, trnc(z^{\leq(i)})) \mid \neg E_i]\right| + 2\mathbb{P}[E_i] \\
&= \left|\mathbb{E}_Z[D(0, trnc(z^{\leq(i+1)})) \mid \neg E_i] - \mathbb{E}_Z[D(0, z^{\leq(i)}) \mid \neg E_i]\right| + 2\mathbb{P}[E_i] \\
&\leq \frac{120\epsilon c^2 (2p)^2}{N^{1/2}} + (2p)^4 N^3 + \frac{8}{N^5} + \frac{4}{N^5}.
\end{aligned}$$

The equality in the fourth line follows from the fact that whenever E_i does not occur, $trnc(z^{\leq(i)}) = z^{\leq(i)}$ by definition. The last inequality follows from inequalities (7) and (8). Along with the fact that $t = \frac{1}{p^2} = 16N^4$, and $\epsilon \leq 1$, this implies that

$$\begin{aligned}
\Delta &\leq \sum_{i=0}^{t-1} \Delta_i \leq t\left(\frac{120\epsilon c^2 (2p)^2}{N^{1/2}} + (2p)^4 N^3 + \frac{12}{N^5}\right) \leq \frac{480\epsilon c^2}{N^{1/2}} + 16p^2 N^3 + \frac{192}{N} \\
&= O\left(\frac{c^2}{N^{1/2}} + \frac{1}{N}\right) = O\left(\frac{c^2}{N^{1/2}}\right).
\end{aligned}$$

The last inequality follows from the assumption that $c \geq 1$. This completes the proof of Theorem 3. ◀

References

1 Scott Aaronson. BQP and the polynomial hierarchy. In *STOC 2010*. ACM, 2010. doi:10.1145/1806689.1806711.

2 Scott Aaronson and Andris Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. In *STOC 2015*. ACM, 2015. doi:10.1145/2746539.2746547.

3 Ziv Bar-Yossef, T. S. Jayram, and Iordanis Kerenidis. Exponential separation of quantum and classical one-way communication complexity. In *STOC 2004*. ACM, 2004. doi:10.1145/1007352.1007379.

4 Harry Buhrman, Richard Cleve, and Avi Wigderson. Quantum vs. classical communication and computation. In *STOC 1998*. ACM, 1998. doi:10.1145/276698.276713.

5 Arkadev Chattopadhyay, Yuval Filmus, Sajin Koroth, Or Meir, and Toniann Pitassi. Query-to-communication lifting for BPP using inner product. In *ICALP 2019*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019. doi:10.4230/LIPIcs.ICALP.2019.35.

6 Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. Pseudorandom generators from polarizing random walks. In *CCC 2018*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPIcs.CCC.2018.1.

7 Eshan Chattopadhyay, Pooya Hatami, Shachar Lovett, and Avishay Tal. Pseudorandom generators from the second fourier level and applications to AC0 with parity gates. In *ITCS 2019*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019. doi:10.4230/LIPIcs.ITCS.2019.22.

8 Dmitry Gavinsky. Entangled simultaneity versus classical interactivity in communication complexity. In *STOC 2016, Cambridge, MA, USA, June 18-21, 2016*. ACM, 2016. doi:10.1145/2897518.2897545.

9 Dmitry Gavinsky, Julia Kempe, Iordanis Kerenidis, Ran Raz, and Ronald de Wolf. Exponential separations for one-way quantum communication complexity, with applications to cryptography. In *STOC 2007*. ACM, 2007. doi:10.1145/1250790.1250866.

10 Mika Göös, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for BPP. In *FOCS 2017*. IEEE Computer Society, 2017. doi:10.1109/FOCS.2017.21.

11 Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Structure of protocols for XOR functions. In *FOCS 2016*. IEEE Computer Society, 2016. doi:10.1109/FOCS.2016.38.

12 Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press 2014, 2014.

13 Ran Raz. Fourier analysis for probabilistic communication complexity. *Comput. Complex.*, 1995. doi:10.1007/BF01206318.

14 Ran Raz. Exponential separation of quantum and classical communication complexity. In *STOC 1999*. ACM, 1999. doi:10.1145/301250.301343.

15 Ran Raz and Avishay Tal. Oracle separation of BQP and PH. In *STOC 2019*. ACM, 2019. doi:10.1145/3313276.3316315.

16 Oded Regev and Bo'az Klartag. Quantum one-way communication can be exponentially stronger than classical communication. In *STOC 2011*. ACM, 2011. doi:10.1145/1993636.1993642.