Entangled simultaneity versus classical interactivity in communication complexity*

Dmitry Gavinsky†

February 19, 2020

Abstract

In 1999 Raz demonstrated a partial function that had an efficient quantum two-way communication protocol but no efficient classical two-way protocol and asked whether there existed a function with an efficient quantum one-way protocol, but still no efficient classical two-way protocol. In 2010 Klartag and Regev demonstrated such a function and asked whether there existed a function with an efficient quantum simultaneous-messages protocol, but still no efficient classical two-way protocol.

In this work we answer the latter question affirmatively and present a partial function Shape that can be computed by a protocol sending entangled simultaneous messages of poly-logarithmic size, and whose classical two-way complexity is lower bounded by a polynomial.

1 Introduction

The setting of communication complexity is one of the strongest computational models where we already have tools to prove “hardness” – that is, where we know problems that don’t admit efficient solutions. In particular, we can compare the computational power of different communication complexity regimes (classes) via demonstrating that certain problem has an efficient solution in one regime, but not in the other.

There are three main types of the problems used for separating communication complexity classes: total functions, partial functions and relations. They form a “hierarchy” in the following sense: separations via total functions are the strongest (“most convincing”), separations via relations are the weakest (and thus the easiest to obtain) and separations via partial functions are “in between”. There are known cases where a quantum communication complexity class can be separated from a classical one via a relation, while a functional separation is provably impossible (cf. [GRdW08]).

The history of (exponential) separations in communication complexity that have demonstrated the advantage of quantum communication can be briefly outlined as follows.

---

*This work was presented at the 48th Symposium on Theory of Computing in 2016.
†Institute of Mathematics, Academy of Sciences, Zitna 25, Praha 1, Czech Republic. Partially funded by the grant P202/12/G061 of GA ČR and by RVO: 67985840. Part of this work was done while visiting the Centre for Quantum Technologies at the National University of Singapore, and was partially funded by the Singapore Ministry of Education and the NRF.

1By calling a problem hard we always mean that no efficient solution exists (this is somewhat different from the usual notion of hardness in the context of computational complexity).
• In 1999 Raz [Raz99] demonstrated a partial function that had an efficient quantum two-way communication protocol, but no efficient classical two-way protocol.
• In 2001 Buhrman, Cleve, Watrous and de Wolf [BCWdW01] showed that a total function (namely, the equality function) had an efficient quantum simultaneous-messages protocol without shared randomness, but no efficient classical simultaneous-messages protocol without shared randomness.
• In 2004 Bar-Yossef, Jayram and Kerenidis [BYJK04] demonstrated a relation that had an efficient quantum one-way communication protocol, but no efficient classical one-way protocol.
• In 2008 in a joint work with Kempe, Kerenidis, Raz and de Wolf [GKK+08] the same separation was demonstrated via a partial function.
• In 2008 a relation was demonstrated [Gav08] with an efficient quantum one-way protocol, but no efficient classical two-way protocol; the next year the result was strengthened [Gav09], to admit an efficient entangled simultaneous-messages protocol.
• In 2010 Klartag and Regev [KR11] demonstrated a partial function with an efficient quantum one-way protocol, but no efficient classical two-way protocol.
• More recently a partial function was demonstrated [Gav19] with an efficient quantum simultaneous-messages protocol without shared randomness, but no efficient classical simultaneous-messages protocol, even with shared randomness.

In this work we demonstrate a partial function that can be solved by a quantum simultaneous-messages protocol that sends $O(\log^2 n)$ entangled qubits; the classical two-way communication complexity of the same function is $\Omega(\sqrt{n})$. In particular, this answers an open question stated by Klartag and Regev.

A number of researchers believe that strong separations of this type are not possible for total functions – that is, unless the model is “too weak” (like simultaneous message passing without shared randomness, used in [BCWdW01]), it might be able to efficiently “emulate” classically any quantum protocol that computes a total function. Proving or disproving this hypothesis is a major open problem. However, if it is true, then the result of this work is very close to an “as strong as possible” demonstration of qualitative advantage of quantum communication over the classical one.

We note that our communication problem also has a quantum one-way protocol of cost $O(\log^2 n)$ (as follows from the existence of a simultaneous-messages protocol of cost $O(\log^2 n)$ that uses $O(\log^2 n)$ bits of entanglement). Till now, the only known example of super-polynomial advantage of quantum one-way over classical two-way communication in solving a functional problem has been the one demonstrated by Klartag and Regev. Their communication problem is called Vector in Subspace Problem (VSP) (originally proposed by Kremer [Kre95]), and it is complete for functions in quantum one-way: that is, the input to any function that admits an efficient quantum one-way protocol can be mapped locally to an instance of VSP that has the same answer as the original problem and solving which is efficient in terms of the original input size. Accordingly, the present work gives an alternative proof of qualitative hardness of VSP for classical two-way communication: if there were an efficient protocol for VSP, there would be one for our problem as well, and we prove the opposite.

\[ ^2 \text{both total and partial, though not for relations!} \]

\[ ^3 \text{Quantitatively, the lower bound given in [KR11] for the classical two-way complexity of VSP is much stronger: it is } \Omega\left(n^{3/5}\right), \text{ while the following argument gives only } 2^{\Omega(\sqrt{\log n})}. \]
A definition of our communication problem and an overview of our approach are given in Sections 3 and 3.1, respectively.

2 Preliminaries

For \( x \in \{0,1\}^n \) and \( 1 \leq i \leq n \), we will use both \( x_i \) and \( x(i) \) to address the \( i \)’th bit of \( x \). Similarly, for \( S \subseteq \{1,\ldots,n\} \), both \( x_S \) and \( x(S) \) will denote the \( |S| \)-bit string, consisting of naturally-ordered bits of \( x \) whose indices are in \( S \). For \( y \in \{0,1\}^n \), let \( x_y \overset{\text{def}}{=} x_{\{i|y_i=1\}} \). Let \(|x|\) denote the Hamming weight of \( x \). Let \( x \oplus y \), \( x \land y \), \( x \lor y \) and \( \lnot x \) denote, respectively, the bit-wise XOR, AND, OR and NOT operations.

For any discrete set \( A \), let \( U_A \) denote the uniform distribution on \( A \). Sometimes (e.g., in subscripts) we will write \( \sim \) instead of \( \sim U_A \). We will sometimes emphasise that a distribution on \( \{0,1\}^2n \) is “viewed as bipartite” (i.e., assumed to be the joint distribution of two random variables, containing \( n \) bits each) by calling it a distribution on \( \{0,1\}^{n+n} \); similarly, we will write \( (X,Y) \in \{0,1\}^{n+n} \), etc.

We will often use the following property of the entropy operator: for any random variables \( X, Y \) and \( Z \) and event \( e \),

\[
H(X|Y,e) \geq H(X|Y,Z,e).
\]

(1)

We will refer to this as the non-growth of entropy under conditioning.\(^5\)

For \( p \in [0,1] \) we will denote by \( h_2(p) \) the binary entropy function, defined as

\[
h_2(p) \overset{\text{def}}{=} p \cdot \log_p 1 - p \cdot \log_p 1 - p
\]

on \((0,1)\), and \( h_2(0), h_2(1) \overset{\text{def}}{=} 0 \). Its Taylor expansion around \( p = 1/2 \) gives

\[
1 - h_2(p) = \frac{1}{2 \ln 2} \cdot \sum_{i=1}^{\infty} \frac{(1 - 2p)^{2i}}{i \cdot (2i - 1)}.
\]

(2)

The following is a weak chain rule for min-entropy, equipped with a tail bound:

**Claim 1.** Let \( \nu \) be a bipartite distribution on the product set \( A \times B \). Then

\[
\mathbb{E}_{(Y_1,Y_2)\sim \nu} \left[ H_{\min} \left( X_2 | (X_1,Y_1) = (Y_1) \right) \right] \geq H_{\min} (\nu) - H \left( X_1 | X_2 \sim \nu \right) (X_1) \geq H_{\min} (\nu) - \log |A|\]

\(^4\)In the case of min-entropy we only allow conditioning on events, defined as the min-entropy of the corresponding distribution.

\(^5\) Note that the entropy operator is monotonically non-increasing with respect to added conditioning on variables, but in general not with respect to conditioning on events.
and for any $\Delta \geq 0$:

$$
\Pr_{(Y_1,Y_2) \sim \nu} \left[ \frac{H_{\min} (X_2 | X_1 = Y_1)}{H_{\min} (\nu)} - \log |A| - \Delta \right] \leq 2^{-\Delta}.
$$

**Proof.** For all $x_1 \in A$ and $x_2 \in B$,

$$
H_{\min} (\nu) \leq \log \frac{1}{\Pr_{\nu} [X_1 = x_1, X_2 = x_2]} = \log \frac{1}{\Pr_{\nu} [X_1 = x_1]} \cdot \frac{1}{\Pr_{\nu} [X_2 = x_2 | X_1 = x_1]} = \log \frac{1}{\Pr_{\nu} [X_1 = x_1]} + \log \frac{1}{\Pr_{\nu} [X_2 = x_2 | X_1 = x_1]}. \tag{3}
$$

Taking $x_2$ that minimises the second term,

$$
H_{\min} (\nu) \leq \log \frac{1}{\Pr_{\nu} [X_1 = x_1]} + H_{\min} (X_2 | X_1 = x_1).
$$

Averaging it over all $x_1$'s gives the first desired inequality:

$$
H_{\min} (\nu) \leq \mathbb{E}_{(Y_1,Y_2) \sim \nu} \left[ \log \frac{1}{\Pr_{\nu} (X_1, X_2) \sim \nu | X_1 = Y_1} + H_{\min} (X_2 | X_1 = Y_1) \right] = H_{(X_1,X_2) \sim \nu} (X_1) + \mathbb{E}_{(Y_1,Y_2) \sim \nu} \left[ H_{\min} (X_2 | X_1 = Y_1) \right].
$$

Averaging (3) over

$$
A_\Delta \overset{\text{def}}{=} \left\{ x_1 \in A \mid H_{\nu} (X_2 | X_1 = x_1) \leq H_{\min} (\nu) - \log |A| - \Delta \right\}
$$

gives the tail bound:

$$
H_{\min} (\nu) \leq \sum_{x_1 \in A_\Delta} \Pr_{\nu} [X_1 = x_1] \cdot \left( \log \frac{1}{\Pr_{\nu} [X_1 = x_1]} + H_{\nu} (X_2 | X_1 = x_1) \right)
$$

$$
= H_{(X_1,X_2) \sim \nu} (X_1 | X_1 \in A_\Delta) + \log \frac{1}{\Pr_{\nu} [X_1 \in A_\Delta]} + \sum_{x_1 \in A_\Delta} \frac{\Pr_{\nu} [X_1 = x_1]}{\Pr_{\nu} [X_1 \in A_\Delta]} \cdot H_{\nu} (X_2 | X_1 = x_1)
$$

$$
\leq \log |A| + \log \frac{1}{\Pr_{\nu} [X_1 \in A_\Delta]} + H_{\min} (\nu) - \log |A| - \Delta,
$$

and so, $\log \frac{1}{\Pr_{\nu} [X_1 \in A_\Delta]} \geq \Delta$, as required. 

We will use the following bound on the $l_1$-distance between two distributions:

**Claim 2.** Let $\nu_1$ and $\nu_2$ be distributions on $\{0,1\}^n$. Then

$$
\|\nu_1 - \nu_2\|_1^2 \leq 8 \ln 2 \cdot \left( n - \min \left\{ H (\nu_1), H (\nu_2) \right\} \right).
$$
Proof. Let $u$ be the uniform distribution on $\{0, 1\}^n$, then

$$H(\mu) = \sum_x \mu(x) \log \frac{1}{\mu(x)} = -\sum_x \mu(x) \log \frac{\mu(x)}{2^{-n}} + \sum_x \mu(x) \log \frac{1}{2^{-n}} = n - d_{KL}(\mu\|u).$$

From the triangle and Pinsker’s inequalities,

$$\|\nu_1 - \nu_2\|_1^2 \leq 4 \max_{j \in \{1, 2\}} \left\{\|\nu_j - u\|_1^2\right\} \leq 8 \ln 2 \cdot \max_{j \in \{1, 2\}} \left\{d_{KL}(\nu_j\|u)\right\} = 8 \ln 2 \cdot \left(\max_{j \in \{1, 2\}} \left\{H(\nu_j)\right\}\right).$$

Let $S_n$ denote the group of permutations of the set $\{1, \ldots, n\}$, and let $\sigma_i \in S_n$ be the $i$'th cyclic permutation (i.e., $\sigma_i(j) = i + j$ if $i + j \leq n$ and $i + j - n$ otherwise). For $x \in \{0, 1\}^n$ and $\tau \in S_n$, denote by $\tau(x)$ the element of $\{0, 1\}^n$ whose $\tau(i)$'th position contains $x_i$ for each $i$ – in particular, $\sigma_j(x)$ is the $j$-bit cyclic shift of $x$.

3 Communication complexity

Please see [KN97] for an extensive overview of classical communication complexity. The quantum counterparts differ from the classical communication models in two aspects: the players are allowed to send quantum messages (accordingly, the complexity is measured in qubits) and to perform arbitrary quantum operations locally. We say that a communication model allows prior entanglement if the players can share any (input-independent) quantum state and use it in the protocol (in the case of simultaneous message passing, entanglement is only allowed between Alice and Bob).

The communication problem that we use for our separation is the following partial function.

**Definition 1 (Shape – Shifted Approximate Equality).** Let Alice receive $(x_1, x_2) \in \{0, 1\}^{n+n}$ and Bob receive $(y_1, y_2) \in \{0, 1\}^{n+n}$. Then

$$\text{Shape}(x_1, x_2, y_1, y_2) = \begin{cases} 1 & \text{if } \exists i : |\sigma_i(x_1) \oplus x_2 \oplus \sigma_i(y_1) \oplus y_2| \leq \frac{2n}{5}; \\ 0 & \text{if } \forall i : \frac{7n}{15} \leq |\sigma_i(x_1) \oplus x_2 \oplus \sigma_i(y_1) \oplus y_2| \leq \frac{8n}{15}; \\ \text{undefined} & \text{otherwise.} \end{cases}$$

That is, $\text{Shape}(x_1, x_2, y_1, y_2)$ “asks” whether there exists a cyclic shift $\sigma_i$, such that $\sigma_i(x_1) \oplus x_2$ is close to $\sigma_i(y_1) \oplus y_2$. Note that both the “meaning of closeness” and the promised gap in the definition of $\text{Shape}$ are $\Omega(n)$ – as we will see, the former is crucial for the lower bound argument, while the latter is (apparently) essential in order to admit an efficient quantum protocol.
3.1 Intuition behind the Shape and its analysis

To design a simultaneous-messages protocol for Shape, we will first build a protocol for the following "sub-problem":

\[ \text{Shape}_i(x_1, x_2, y_1, y_2) \overset{\text{def}}{=} \begin{cases} 1 & \text{if } |\sigma_i(x_1) \oplus x_2 \oplus \sigma_i(y_1) \oplus y_2| \leq \frac{2n}{3}; \\ 0 & \text{if } \frac{7n}{15} \leq |\sigma_i(x_1) \oplus x_2 \oplus \sigma_i(y_1) \oplus y_2| \leq \frac{8n}{15}; \\ \text{undefined} & \text{otherwise.} \end{cases} \]

Our protocol for Shape will be such that the quantum messages sent by the players will be independent of i, and only the referee will need to know i in order to measure the messages and produce the answer. Accordingly, reducing the error of solving Shape to sufficiently small (inverse-polynomial) value and sequentially applying the referee’s measurements corresponding to all 0 \leq i < n to the same quantum messages received from the players will produce answers to all n instances of Shape, (with respect to the actual input), and with high probability all these answers will be correct. This gives a protocol for Shape.

To show hardness of Shape for classical two-way communication, we start by using relatively standard entropy-based arguments to say – towards contradiction – that if there is a short protocol for Shape, then there exists a large subset \( A \subseteq \{0, 1\}^{n+1} \), such that when \( (X, Y) \sim \mathcal{U}_A \), “something” is known about \( \sigma_i(X) \oplus Y \) for every 0 \leq i < n, that is, \( \Omega(1) \) bits of entropy are “missing”, on average, in each of the corresponding distributions. One possible way for \( A \) to have this property would be to “fix” certain bits of all its elements, and the number of fixed bit-positions should be sufficient in order to have them “overlap” in every \( \sigma_i(x) \oplus y \) when \( (x, y) \in A \) – that is, roughly \( \Omega(\sqrt{n}) \) positions must be fixed (according to the “birthday paradox”), which means that \( A \) can have size at most \( 2^{2n-\Omega(\sqrt{n})} \), which, in turn, is sufficient for our lower bound.

However, there is another possibility for \( A \) to have the same property: it can to fix only the bit-parity of all its elements and have size \( 2^{2n-1} \), thus leading to no meaningful lower bound. This is the reason why Shape has been defined “with margins” (the problem asks whether the bit-wise XOR of strings is close to 0 – not necessarily equals it): this way we can choose the input distribution to be “noisy” and draw a stronger conclusion from the existence of a short protocol, namely that “something” is known about every \( \sigma_i(X) \oplus \tilde{Y} \) when \( (X, Y) \sim \mathcal{U}_A \), where \( \tilde{Y} \) is a noisy version of the random variable \( Y \). Noise is known to “damage” high-degree Fourier coefficients (as quantified by the famous hypercontractive inequality) – in particular, a set \( A \) that only restricts the bit-parity of its elements would fail miserably with respect to the “noisy condition”.

We would like to use the hypercontractivity to prove an upper bound on the size of \( A \); indeed, the “noisy condition” is very similar to what has been analysed in [GKK+08] and led to a good bound on \( |A| \). There is an interesting distinction between the two cases: in [GKK+08] “something” had to be known about \( \tau(X) \oplus \tilde{Y} \) for every \( \tau \in S_n \), whereas in this work we can only require that it holds for the \( \sigma_i \)'s.\(^6\) The argument in [GKK+08] relies strongly on the symmetry resulting from allowing all n! permutation – our modestly-permuted Shape seems to ask for different treatment.

\(^6\)Instead of \( \{\sigma_i|0 \leq i < n\} \) we could have fixed any other family, containing up to quasi-polynomial number of permutations; it may not be larger than that, as our quantum protocol would no longer be able to solve the problem efficiently.
There are at least two naturally-looking approaches to analyse a set \( A \) that satisfies the “noisy condition” with respect to a small family of permutations. First, we can use the fact that any such family corresponds to a rather small (of size roughly \( 2^n \cdot \{\text{the number of permutations}\} \)) subset of all \( 2^{\binom{n}{2}} \) Fourier coefficients in the characteristic function of \( A \) which are responsible for the entire entropy loss – we can try to investigate the Fourier structure of Boolean functions that are “heavily supported” on the corresponding subset. Second, we can use entropy-inspired arguments and try to show that staying distinguishable after the action of one of polynomially-many allowed permutations followed by noise is, essentially, as hard as staying distinguishable after the action of a uniformly-random (disclosed) permutation from \( S_n \). The proof will use the second approach.\(^7\)

4 Solving Shape with simultaneous entangled messages

Here we give a protocol for solving Shape in \( Q^{\|\text{ent}} \) – the model of quantum simultaneous message passing with entanglement (between Alice and Bob).

Protocol for Shape:\quad In the beginning Alice and Bob share two states: \( |A_0\rangle = \frac{1}{\sqrt{n}} \cdot \sum_{k=1}^{n} |k\rangle|k\rangle \) and \( |B_0\rangle = \frac{1}{\sqrt{n}} \cdot \sum_{j=1}^{n} |j\rangle|j\rangle \), where Alice holds the first “register” and Bob holds the second “register” of each state. Upon receiving the input, the players apply local conditional phase-negations to transform the shared states into the form

\[
|A_1\rangle = \frac{1}{\sqrt{n}} \cdot \sum_{k=1}^{n} (-1)^{x_1(k) + y_1(k)} |k\rangle|k\rangle, \quad |B_1\rangle = \frac{1}{\sqrt{n}} \cdot \sum_{j=1}^{n} (-1)^{x_2(j) + y_2(j)} |j\rangle|j\rangle.
\]

Then the players send the parts of the shared states to the referee who applies \( \sigma_{-i} \) to both registers of \( |A_1\rangle \), resulting in

\[
|A_2\rangle = \frac{1}{\sqrt{n}} \cdot \sum_{k=1}^{n} (-1)^{x_1(k) + y_1(k)} |\sigma_{-i}(k)\rangle|\sigma_{-i}(k)\rangle = \frac{1}{\sqrt{n}} \cdot \sum_{k=1}^{n} (-1)^{x_1(\sigma_i(k)) + y_1(\sigma_i(k))} |k\rangle|k\rangle.
\]

At this point,

\[
\langle A_2 | B_1 \rangle = \frac{1}{n} \cdot \sum_{k=1}^{n} (-1)^{x_1(\sigma_i(k)) + x_2(k) + y_1(\sigma_i(k)) + y_2(k)} = 1 - \frac{2}{n} \cdot |\sigma_i(x_1) \oplus x_2 \oplus \sigma_i(y_1) \oplus y_2|
\]

– in particular,

\[
|\langle A_2 | B_1 \rangle| \begin{cases} \leq \frac{1}{15} & \text{if} \ Shape_i(x_1, x_2, y_1, y_2) = 0, \\ \geq \frac{1}{5} & \text{if} \ Shape_i(x_1, x_2, y_1, y_2) = 1. \end{cases}
\]

The referee can distinguish the two cases with confidence \( \frac{1}{2} + \Omega(1) \) by performing the swap-test – a two-outcome measurement that “accepts” with probability \( \frac{1+|\langle \phi_1 | \phi_2 \rangle|^2}{2} \) when performed over the pair of (pure) quantum states \( |\phi_1\rangle \) and \( |\phi_2\rangle \).

\(^7\)We still use the hypercontractivity to conclude the proof, in a way somewhat similar to [KLL88] (cf. Claim 3); in the language of the above informal description, that is done after using the presence of noise in the input distribution in order to limit (via entropy-inspired arguments) the potential role of high-degree Fourier coefficients.
The above protocol can be repeated $O(\log \frac{1}{\varepsilon})$ times in parallel to bring the error down to any $\varepsilon > 0$ – let $\mathcal{P}_{i,\varepsilon}$ denote the resulting protocol. The total communication cost of $\mathcal{P}_{i,\varepsilon}$ is $O(\log n \cdot \log \frac{1}{\varepsilon})$ and it uses $O(\log n \cdot \log \frac{1}{\varepsilon})$ bits of entanglement.

Let $(\Pi_{i,\varepsilon}, I - \Pi_{i,\varepsilon})$ be the 2-outcome projective measurement that the referee applies in $\mathcal{P}_{i,\varepsilon}$ to the messages received from the players in order to determine the answer (with outcome $\Pi_{i,\varepsilon}$ corresponding to answering “Shape$(x_1, x_2, y_1, y_2) = 1$”), and let this be the only step performed by the referee.\(^8\) Note that running $\mathcal{P}_{i,\varepsilon}$ didn’t require either Alice or Bob to know the actual value of $i$ – only the referee had to know it in order to apply $(\Pi_{i,\varepsilon}, I - \Pi_{i,\varepsilon})$. This makes $\mathcal{P}_{i,\varepsilon}$ a perfect “building block” for solving the original problem.

**Protocol for Shape:** Let Alice and Bob send to the referee their messages, as prescribed by $\mathcal{P}_{1,\varepsilon'}$, for some $\varepsilon'$ to be fixed soon. The referee sequentially measures the received messages with $(\Pi_{i,\varepsilon'}, I - \Pi_{i,\varepsilon'})$ for all $0 \leq i < n$. If at least one outcome $\Pi_{i,\varepsilon'}$ has been obtained, the referee answers “Shape$(x_1, x_2, y_1, y_2) = 1$”; otherwise, “Shape$(x_1, x_2, y_1, y_2) = 0$”.

Call the above protocol $\mathcal{P}$. Assume without loss of generality that $\text{Shape}(x_1, x_2, y_1, y_2) \in \{0, 1\}$ (i.e., the input fulfills the promise). To analyse the error of $\mathcal{P}$, note that the protocol can return the wrong answer only if at some round $i$ the outcome of the measurement $(\Pi_{i,\varepsilon'}, I - \Pi_{i,\varepsilon'})$ was wrong – that is, the outcome was $\Pi_{i,\varepsilon'}$ while $\text{Shape}_i(x_1, x_2, y_1, y_2) = 0$, or vice versa. Note that while the probability of the outcome of the first performed measurement being wrong is bounded above by $\varepsilon'$ (as follows trivially from the error bound of $\mathcal{P}_{1,\varepsilon'}$), at the subsequent rounds the state being measured may have been “distorted” by the earlier measurements, which, in turn, may increase the error probability.

We analyse\(^9\) the probability that the first wrong outcome during execution of $\mathcal{P}$ has occurred at round $j$ – denote it by $\varepsilon_j$. For all $0 \leq i < j$, let $\Pi_{i,\varepsilon'} \in \{\Pi_{i,\varepsilon'}, I - \Pi_{i,\varepsilon'}\}$ be the right outcome for round $i$. Let $v$ be a unit vector representing the (pure) quantum state of the original messages received by the referee from the players (before any measurement has been applied). Denote:

$$
\Pi_{i,\varepsilon'} v = v + u_i.
$$

From correctness of $\mathcal{P}_{1,\varepsilon'}$s it follows that $\|u_i\|_2 \leq \sqrt{\varepsilon'}$.

Let $v_j$ be the input state for the $j$th measurement. As we are assuming that all the previous measurements produced the right answers,

$$
v_j = \Pi_{j-1,\varepsilon'} \cdots \Pi_{1,\varepsilon'} v = \frac{v + \Pi_{j-1,\varepsilon'} u_1 + \cdots + \Pi_{j-1,\varepsilon'} u_{j-2}}{\|\Pi_{j-1,\varepsilon'} \cdots \Pi_{1,\varepsilon'} v\|_2} = \alpha(v + v_j'),
$$

where $\alpha = 1/\|\Pi_{j-1,\varepsilon'} v\|_2 \geq 1$ and $\|v_j'\|_2 \leq \sum_{i=1}^{j-1} \|u_i\|_2 < n\sqrt{\varepsilon'}$. Then

$$
\varepsilon_j = 1 - \|\Pi_{j,\varepsilon'} v_j\|_2 = 1 - \alpha^2 \|\Pi_{j,\varepsilon'} v + \Pi_{j,\varepsilon'} v_j'\|_2^2 \leq 1 - \|v \cdot (1 - \|u_j\|_2 - \|v_j'\|_2\|_2^2 < 4n\sqrt{\varepsilon'}.
$$

\(^8\)That is, the measurement $(\Pi_{i,\varepsilon'}, I - \Pi_{i,\varepsilon'})$ “contains” all the steps taken by the referee according to $\mathcal{P}_{i,\varepsilon}$.

Alternatively, we can say that the referee “uncomputes” his previous (recursive) steps upon having performed the (only) measurement in the end of $\mathcal{P}_{i,\varepsilon}$.

\(^9\)Below we give a simple proof of a sub-optimal bound, sufficient for us. A more delicate treatment of similar setting is implicit in [ANTSV02] and explicit in Lemma 2 of [Aar04].
The error probability of \( \mathcal{P} \) is at most \( \sum_{j=1}^{n} \varepsilon_j < 4n^2 \sqrt{\varepsilon'} \) – choosing \( \varepsilon' = \frac{\varepsilon^2}{16n^2} \) makes it less than \( \varepsilon \). The resulting communication cost and entanglement requirements of \( \mathcal{P} \) are \( O(\log n \cdot \log \frac{1}{\varepsilon'}) = O(\log^2 n + \log n \cdot \log \frac{1}{\varepsilon'}) \).

**Theorem 1.** The complexity of \( \text{Shape} \) in \( \mathcal{Q}^{\parallel,\text{ent}} \) is \( O(\log^2 n) \).

## 5 Solving \text{Shape} with classical interaction

In this and the next sections we prove a lower bound on the complexity of \( \text{Shape} \) in \( \mathcal{R} \) – the model of interactive classical communication:

**Theorem 2.** The complexity of \( \text{Shape} \) in \( \mathcal{R} \) is \( \Omega(\sqrt{n}) \).

The above bound is nearly-tight.\(^{10}\)

We start by introducing several distributions.

**Definition 2 (Useful distributions – \( T_\delta, \mu_0, \mu_1 \) and \( \mu \)).**

- For \( \delta \in [0,\frac{1}{2}] \), let \( T_\delta \) be the distribution on \( \{0,1\}^n \), where each position is 1 with probability \( \delta \), independently from the other positions. For a random variable \( X \sim \nu \) taking values from \( \{0,1\}^n \), let \( T_\delta(X) \) be the variable \( X \oplus Z \) and \( T_\delta(\nu) \) be the distribution of \( T_\delta(X) \), and for \( x \in \{0,1\}^n \), let \( T_\delta(x) \) be the variable \( x \oplus Z \), where \( Z \sim T_\delta \) in all cases.
- Let \( \mu_0 \) be the uniform distribution on \( \{0,1\}^{4n} \).
- For \( 0 \leq i < n \), let \( \mu_1^{(i)} \) be the uniform distribution of \( (X_1, X_2, Y_1, Y_2) \) taking values from \( \{0,1\}^{4n} \), modulo the condition \( \sigma_1(X_1) \oplus X_2 \oplus \sigma_1(Y_1) \oplus Y_2 = 0 \), and let \( \tilde{\mu}_1^{(i)} \) be the distribution of \( (T_{j/8}(X_1), X_2, Y_1, Y_2) \) when \( (X_1, X_2, Y_1, Y_2) \sim \mu_1^{(i)} \). Let \( \mu \) be the distribution of sampling from \( \tilde{\mu}_1^{(i)} \) for uniformly random \( i \).
- Let \( \mu \) be the distribution of sampling \( (X_1, X_2, Y_1, Y_2) \sim \mu_Z \) when \( Z \sim \mathcal{U}_{\{0,1\}} \).

Note that

\[
\forall j \in \{0,1\} : \mathbf{Pr}_{\mu_j} [\text{Shape}(X_1, X_2, Y_1, Y_2) = j] \in 1 - o(1),
\]

and therefore,

\[
\mathbf{Pr}_\mu [\text{Shape}(X_1, X_2, Y_1, Y_2) = j] \in \frac{1}{2} \pm o(1).
\]

We will show that solving \( \text{Shape} \) under \( \mu \) with small constant error\(^{11}\) requires \( \Omega(\sqrt{n}) \) bits of communication.

\(^{10}\)Due to the “birthday paradox”, a single message from Alice that contains the values of \( \Theta(\sqrt{n \log n}) \) randomly-chosen bits of \( (x_1, x_2) \) allows Bob to answer \( \text{Shape}(x_1, x_2, y_1, y_2) \) with polynomially-small error probability – using shared randomness to choose the positions, this can be implemented by a protocol of cost \( O(\sqrt{n \log n}) \). A similar approach can be used to get an \( O(\sqrt{n \log n}) \)-bit protocol in \( \mathcal{R}^{\parallel,\text{pub}} \) – the model of classical simultaneous message passing with shared randomness.

\(^{11}\)The support of \( \mu \) is \( \{0,1\}^{4n} \) – in particular, sometimes \( \text{Shape} \) is undefined under \( \mu \); however, this happens with probability \( o(1) \) and a good protocol must return the right answer from \( \{0,1\} \) for the given input always, except with arbitrarily small constant probability.
Proof outline

Lower-bounding the $\mathcal{R}$-complexity of a communication problem is often based on showing that it admits no large nearly-monochromatic rectangles with respect to certain input distribution: here a (combinatorial) rectangle is a subset of input pairs of the product form $A \times B$, where $A$ and $B$ are, respectively, subsets of Alice’s and Bob’s inputs. We use this approach in the analysis of Shape.

- In Step 1 we show that if a short protocol were solving Shape, then there would exist a large input rectangle $A \times B$, strongly biased towards “Shape($x_1, x_2, y_1, y_2$) = 0” under $\mu$.
- In Step 2 we say that if the above were true, then either $A$ or $B$ would “tell something” about each possible “noisy shifted convolution” of its elements (i.e., $\Omega(1)$ bits would be known about every $\sigma_i(x_1) \oplus T_{j,i}(x_2)$, $0 \leq i \leq n - 1$ when $(x_1, x_2)$ is a uniformly-random element of the set).
- In Step 3 we say that if the above were true, then $\Omega(\sqrt{n})$ bits would be known about a uniformly-random element of either $A$ or $B$.
- In Step 4 we conclude that Theorem 2 holds.

Step 1

If there were a short protocol solving Shape with small error, then there would exist a large input rectangle, strongly biased towards “Shape($x_1, x_2, y_1, y_2$) = 0” under $\mu$. Formally:

**Lemma 1.** Assume that a protocol of cost $c$ solves Shape in $\mathcal{R}$ with error at most $\varepsilon$ for some $\varepsilon \in \Omega(1)$. Then for $n$ large enough, there exists a rectangle $A \times B \subseteq \{0, 1\}^{2n} \times \{0, 1\}^{2n}$ of size at least $2^{4n-c-3}$, such that

$$
\mu_1(A \times B) \leq 4\varepsilon \cdot \mu_0(A \times B).
$$

**Proof.** If there is a randomised protocol of cost $c$ that solves Shape with error at most $\varepsilon$, then some value of protocol’s random string achieves at most the same error under the input distribution $\mu$ – by fixing that value we obtain a deterministic protocol $\Pi$ that solves Shape under $\mu$ with error at most $\varepsilon$, while partitioning the set of possible input pairs $((x_1, x_2), (y_1, y_2))$ into (at most) $2^c$ combinatorial rectangles, each labelled with the answer that $\Pi$ gives to the corresponding input pairs. As $\mu$ is the balanced mixture of $\mu_0$ and $\mu_1$, the error of $\Pi$ with respect to both $\mu_0$ and $\mu_1$ is at most $2\varepsilon$.

Let $(X_1, X_2, Y_1, Y_2)$ be random variables taking the corresponding input values. Let $e_1$ be the event that $(X_1, X_2, Y_1, Y_2)$ belongs to a $\Pi$’s rectangle of size at least $2^{4n-c-3}$. As there are at most $2^c$ rectangles and $\mu_0$ is uniform,

$$
\text{Pr}_{\mu_0}[e_1] \geq 7/8.
$$

Let $e_2$ be the event that $(X_1, X_2, Y_1, Y_2)$ belongs to a $\Pi$’s rectangle $A \times B$, such that

$$
\mu_1(A \times B) > 4\varepsilon \cdot \mu_0(A \times B),
$$

and let $q \overset{\text{def}}{=} \text{Pr}_{\mu_0}[\Pi \text{ answers } \text{“}0\text{”} | e_2]$. Then

$$
4\varepsilon \cdot q \cdot \mu_0(e_2) = 4\varepsilon \cdot \text{Pr}_{\mu_0}[e_2 \land \Pi \text{ answers } \text{“}0\text{”}] \leq \text{Pr}_{\mu_1}[e_2 \land \Pi \text{ answers } \text{“}0\text{”}] \leq \text{Pr}_{\mu_1}[\Pi \text{ answers } \text{“}0\text{”}]
$$

10
\[ \leq \Pr_{\mu_1}[\Pi \text{ errs}] + \Pr_{\mu_1}[\text{Shape}(X_1, X_2, Y_1, Y_2) = 0] \in 2\varepsilon + o(1), \]

where the rightmost upper bound is due to (4). Similarly,

\[
(1 - q) \cdot \mu_0(e_2) = \Pr_{\mu_0}[e_2 \land \Pi \text{ answers } "1"] \leq \Pr_{\mu_0}[\Pi \text{ answers } "1"] \\
\leq \Pr_{\mu_0}[\Pi \text{ errs}] + \Pr_{\mu_0}[\text{Shape}(X_1, X_2, Y_1, Y_2) = 1] \in 2\varepsilon + o(1).
\]

From the above two bounds,

\[ 4\varepsilon \cdot \mu_0(e_2) \in 2\varepsilon + o(1) \implies \mu_0(e_2) \in 1/2 + o(1/\varepsilon), \]

and therefore,

\[ \Pr_{\mu_0}[e_1 \land \neg e_2] \in 3/8 - o(1/\varepsilon) \subseteq \Omega(1). \]

The result follows. \hfill \blacksquare

**Step 2**

If input rectangle \( A \times B \) is strongly biased towards “Shape(\( X_1, X_2, Y_1, Y_2 = 0 \)”, then in expectation with respect to uniformly-random \( i, \Omega(1) \) bits are “known” either about \( \sigma_i(X_1) \oplus T_{1/4}(X_2) \) when \( (X_1, X_2) \sim \mathcal{U}_1 \) or about \( \sigma_i(Y_1) \oplus T_{1/4}(Y_2) \) when \( (Y_1, Y_2) \sim \mathcal{U}_B \). Formally:

**Lemma 2.** Let \( \mu_0 \) and \( \mu_1 \) be distributions, as defined in Definition 2. If for some \( \varepsilon \in [0, 1] \) the rectangle \( A \times B \subseteq \{0, 1\}^{2n} \times \{0, 1\}^{2n} \) satisfies \( \mu_1(A \times B) \leq 4\varepsilon \cdot \mu_0(A \times B) \), then

\[
\min \left\{ E_i \left[ H_{(X_1, X_2) \in A} \left( \sigma_i(X_1) \oplus T_{1/4}(X_2) \right) \right], E_i \left[ H_{(Y_1, Y_2) \in B} \left( \sigma_i(Y_1) \oplus T_{1/4}(Y_2) \right) \right] \right\} \leq n - \frac{1}{92} + \varepsilon,
\]

where \( i \sim \mathcal{U}_{\{0, \ldots, n-1\}} \).

**Proof.** By definition, \( \mu_1 = \frac{1}{n} \sum_{i=0}^{n-1} \hat{\mu}_1^{(i)} \), and therefore,

\[ \frac{1}{n} \cdot \sum_{i=0}^{n-1} \hat{\mu}_1^{(i)}(A \times B) \leq 4\varepsilon \cdot \mu_0(A \times B). \]

In particular, for some \( I_0 \subseteq \{0, \ldots, n-1\}, |I_0| \geq n/2: \)

\[ \forall i \in I_0 : \hat{\mu}_1^{(i)}(A \times B) \leq 8\varepsilon \cdot \mu_0(A \times B). \]

Let \( i_0 \in I_0 \). Note that the noise operator is symmetric: \( \Pr[T_\delta(a) = b] = \Pr[T_\delta(b) = a] \) for every \( a \) and \( b \). Furthermore, \( T_\delta(a \oplus b) = T_\delta(a) \oplus b \) and \( T_{\delta_1}(T_{\delta_2}(a)) = T_{\delta_1(\delta_2) + \delta_2(1/4)}(a) \). Therefore,

\[ \hat{\mu}_1^{(i_0)}(x_1, x_2, y_1, y_2) \equiv \frac{\Pr[\sigma_{i_0}(x_1) \oplus T_{1/4}(x_2) = \sigma_{i_0}(y_1) \oplus T_{1/4}(y_2)]}{2^{3n}}. \]
Accordingly,
\[
\frac{8\varepsilon \cdot |A \times B|}{2^{3n}} = 8\varepsilon \cdot \mu_0(A \times B)
\]
\[
\geq \tilde{\mu}^{(i)}_1(A \times B) = \sum_{(x_1, x_2) \in A} \sum_{(y_1, y_2) \in B} \Pr \left[ \sigma_{i_0}(x_1) \oplus T_{1/4}(x_2) = \sigma_{i_0}(y_1) \oplus T_{1/4}(y_2) \right]
\]
and so,
\[
E_{(X_1, X_2) \sim \mathcal{U}_A, (Y_1, Y_2) \sim \mathcal{U}_B} \left[ \Pr \left[ \sigma_{i_0}(X_1) \oplus T_{1/4}(X_2) = \sigma_{i_0}(Y_1) \oplus T_{1/4}(Y_2) \right] \right] \leq \frac{8\varepsilon}{2^n}.
\]

Let \(Z_1 = \sigma_{i_0}(X_1) \oplus T_{1/4}(X_2)\) when \((X_1, X_2) \sim \mathcal{U}_A\) and \(Z_2 = \sigma_{i_0}(Y_1) \oplus T_{1/4}(Y_2)\) when \((Y_1, Y_2) \sim \mathcal{U}_B\). Denote by \(\nu_1\) and \(\nu_2\) the distributions of \(Z_1\) and \(Z_2\), respectively. Then
\[
\frac{8\varepsilon}{2^n} \geq \Pr_{Z_1 \sim \nu_1, Z_2 \sim \nu_2} [Z_1 = Z_2] = E_{Z \sim \nu_1} [\nu_2(Z)].
\]

Therefore, \(\Pr_{Z \sim \nu_1} [\nu_2(Z) \leq 16\varepsilon/2^n] \geq 1/2; \) let \(Z_0 = \{ z \in \{0, 1\}^n \mid \nu_2(z) \leq 16\varepsilon/2^n \}\). Then
\[
\| \nu_1 - \nu_2 \|_1 \geq \sum_{z \in Z_0} |\nu_1(z) - \nu_2(z)| \geq \nu_1(Z_0) - \nu_2(Z_0) \geq \frac{1}{2} - 16\varepsilon.
\]

By Claim 2,
\[
n - \min \left\{ H(\nu_1), H(\nu_2) \right\} \geq \left( \frac{1}{2} - 16\varepsilon \right)^2 \geq \frac{1}{23} - 3\varepsilon.
\]

Finally, since \(H(\sigma_i(X_1) \oplus T_{1/4}(X_2)), H(\sigma_i(Y_1) \oplus T_{1/4}(Y_2)) \leq n\) always,
\[
\sum_{i=0}^{n-1} \left( n - \min \left\{ H_{\mathcal{U}_A} \left( \sigma_i(X_1) \oplus T_{1/4}(X_2) \right), H_{\mathcal{U}_B} \left( \sigma_i(Y_1) \oplus T_{1/4}(Y_2) \right) \right\} \right)
\]
\[
\geq \sum_{i=0}^{n-1} \left( n - \min \left\{ H_{\mathcal{U}_A} \left( \sigma_i(X_1) \oplus T_{1/4}(X_2) \right), H_{\mathcal{U}_B} \left( \sigma_i(Y_1) \oplus T_{1/4}(Y_2) \right) \right\} \right)
\]
\[
> \frac{n}{2} \left( \frac{1}{23} - 3\varepsilon \right)
\]
and so,
\[
\min \left\{ E_{\mathcal{U}_A} \left[ H \left( \sigma_i(X_1) \oplus T_{1/4}(X_2) \right) \right], E_{\mathcal{U}_B} \left[ H \left( \sigma_i(Y_1) \oplus T_{1/4}(Y_2) \right) \right] \right\} < n - \frac{1}{92} + \varepsilon,
\]
as required. \(\blacksquare\) Lemma 2
Step 3
If \( \Omega(1) \) bits are "known" on average about each of \( \sigma_i(X_1) \oplus T_{i/4}(X_2) \) when \( (X_1, X_2) \sim \nu \), then \( \Omega(\sqrt{n}) \) bits are "known" about \( (X_1, X_2) \) itself. Formally:

Lemma 3. If \( \nu \) is a distribution on \( \{0, 1\}^{n+n} \) for large enough \( n \), such that

\[
E_{i \in \{0, \ldots, n-1\}} \left[ H_{(X_1, X_2) \sim \nu} \left( \sigma_i(X_1) \oplus T_{i/4}(X_2) \right) \right] \leq n - \delta
\]

for some \( \delta > 0 \), then

\[
H_{\text{min}}(\nu) \leq 2n - \frac{\sqrt{\delta n}}{29}.
\]

A proof of the lemma is given in Section 6. Note that the noise operator \( T \) is crucial for the statement to hold; without it, \( \nu \) being the uniform distribution over all strings of even parity would provide a counterexample (this has been discussed in more detail in Section 3.1).

Note also that the bound in the conclusion is asymptotically tight, as "entropy deficit" of \( O(\sqrt{n}) \) bits is sufficient for \( \nu \) to satisfy the condition (Sect. 3.1). On the other hand, the proof will not rely on any specific properties of the \( \sigma_i \)'s, and the condition can be somewhat relaxed by assuming sufficient entropy loss on the edge-wise XORs when averaging over a sufficiently large family of pairwise-disjoint perfect matchings between the bits of \( X_1 \) and \( T_{\Omega(1)}(X_2) \).

Step 4
If a protocol of cost \( c \) solves \textit{Shape} in \( \mathcal{R} \) with error at most \( \varepsilon \leq 1/93 \), then Lemma 1 guarantees that the conditions of Lemma 2 are satisfied by some rectangle \( A \times B \) of size at least \( 2^{4n-c-3} \). Lemma 2 guarantees that either \( \nu = \mathcal{U}_A \) or \( \nu = \mathcal{U}_B \) satisfies the condition of Lemma 3 for \( \delta = \frac{1}{92} - \varepsilon \), which, in turn, implies that \( \min(|A|, |B|) \leq 2^{2n-\Omega(\sqrt{n})} \). Theorem 2 follows.

6 Chasing the lost entropy
To prove Lemma 3, we will use some tools related to Boolean Fourier transform and hypercontractivity.\(^{12}\)

Note that even though the noise operator \( T \) is a crucial part of the lemma condition (as discussed earlier), the proof will “handle” \( T \) with information-theoretic methods (cf. Lemma 4), while hypercontractivity (a technique, closely associated with the noise operator) will be used as an analytic tool for analysing the Fourier spectrum of certain “noiseless” probability distribution (cf. Lemma 5).

For \( f, g : \{0, 1\}^n \rightarrow \mathbb{R} \) and \( p > 0 \), let \( \|f\|_p \overset{\text{def}}{=} (E_{x \in \{0, 1\}^n} |f(x)|^p)^{1/p} \), \( \|f\|_{\infty} \overset{\text{def}}{=} \max_x \{|f(x)|\} \) and \( (f, g) \overset{\text{def}}{=} E_x[f(x) \cdot g(x)] \), so that \( \|f\|_2^2 = \langle f, f \rangle \). For \( s \subseteq \{1, \ldots, n\} \) and \( x \in \{0, 1\}^n \), let \( \chi_s(x) \overset{\text{def}}{=} (-1)^{|s|} \) and \( \hat{f}(s) \overset{\text{def}}{=} \langle f, \chi_s \rangle \). The linear mapping \( f \rightarrow \hat{f} \) is \textit{norm-preserving} in the following sense: \( \|f\|_2^2 = \sum_s \hat{f}(s)^2 \) (Parseval’s identity).

\(^{12}\)Note that we are changing the meaning of the notation \( \|v\|_p \); in Sections 4 and 5 it stood for \( (\sum_x |v(x)|^p)^{1/p} \); now we define it to be \( (E_x [|v(x)|^p])^{1/p} \), which is more common in the context of Fourier transform.
Fact 1 (Hypercontractive inequality [Bon70, Bec75]). If $1 \leq p \leq q$, then

$$\left\| \sum_{s \subseteq \{1, \ldots, n\}} \left( \frac{p-1}{q-1} \right)^{|s|/2} \hat{f}(s) \cdot \chi_s \right\|_q \leq \|f\|_p.$$  

From it we derive the following variations on the “KKL theme” (cf. [KKL88]).

Claim 3. Let $f : \{0,1\}^n \to \mathbb{R}$, $\alpha \overset{\text{def}}{=} \|f(x)\|_\infty$ and $\beta \overset{\text{def}}{=} \|f(x)\|_1$. Then for any $\delta \in [0, 1]$:

$$\sum_{s \subseteq \{1, \ldots, n\}} \delta^{\lvert s \rvert} \cdot \hat{f}(s)^2 \leq \alpha^2 \cdot \left( \frac{\beta}{\alpha} \right)^{\frac{2}{1+\delta}},$$

and for any $t \leq 2 \ln \frac{\alpha}{\beta}$:

$$\sum_{\lvert s \rvert \leq t} \hat{f}(s)^2 \leq \beta^2 \cdot \left( \frac{2e \cdot \ln \frac{\alpha}{\beta}}{t} \right)^t.$$  

Proof. Choosing $q = 2$ and $p = 1 + \delta$ in the hypercontractive inequality (Fact 1), we get the first desired bound:

$$\sum_{s} \delta^{\lvert s \rvert} \cdot \hat{f}(s)^2 = \left\| \sum_{s} \delta^{\lvert s \rvert/2} \cdot \hat{f}(s) \cdot \chi_s \right\|_2^2 \leq \|f\|_1^{2+\delta} = \left( \mathbb{E}_x \left[ \|f(x)\|_1^{1+\delta} \right] \right)^{\frac{2}{1+\delta}} = \left( \alpha^{1+\delta} \cdot \mathbb{E}_x \left[ \|f(x)/\alpha\|_1^{1+\delta} \right] \right)^{\frac{2}{1+\delta}} \leq \left( \alpha^{1+\delta} \cdot \mathbb{E}_x \left[ \|f(x)/\alpha\| \right] \right)^{\frac{2}{1+\delta}} = \alpha^{\frac{2\delta}{1+\delta}} \cdot \beta^2,$$

where the first equality follows from Parseval’s identity.

Now observe that

$$\sum_{\lvert s \rvert \leq t} \hat{f}(s)^2 \leq \sum_{s \subseteq \{1, \ldots, n\}} \delta^{\lvert s \rvert} \cdot \hat{f}(s)^2 \leq \alpha^2 \cdot \left( \frac{\beta}{\alpha} \right)^{\frac{2}{1+\delta}} \leq \alpha^2 \cdot \left( \frac{\beta}{\alpha} \right)^{2-2\delta} = \beta^2 \cdot \left( \frac{\alpha}{\beta} \right)^{2\delta},$$

where the second inequality is what we have just proved. Let $\delta \overset{\text{def}}{=} \frac{t}{2 \ln \alpha / \beta}$, then

$$\sum_{\lvert s \rvert \leq t} \hat{f}(s)^2 \leq \frac{\beta^2 \cdot \left( \frac{\alpha}{\beta} \right)^{2\delta}}{\delta^t} = \beta^2 \cdot \left( \frac{\alpha}{\beta} \right)^{\frac{2\delta}{\ln \alpha / \beta}} \left( \frac{2e \cdot \ln \frac{\alpha}{\beta}}{t} \right)^t = \beta^2 \cdot \left( \frac{2e \cdot \ln \frac{\alpha}{\beta}}{t} \right)^t,$$

as required.  

Claim 3

In the rest of this section we prove Lemma 3.

Proof outline

Let us say, informally, that a distribution of bits is noise-resistant if its “noisy version” is distinguishable from the uniform distribution.

14
In Step 1 we observe that a “projection” of a noise-resistant distribution to a randomly-selected subset of bits can (also) be distinguished from the uniform – that is, it cannot have the full entropy. This argument is applied to the distribution of $\sigma_t(X_1) \oplus X_2$, which is noise-resistant by assumption, letting us conclude that a “typical” $(\sigma_t(X_1) \oplus X_2)_S$ has entropy $|S| - \Omega(\delta)$.

In Step 2 we use chain decomposition over $H((\sigma_t(X_1) \oplus X_2)_S)$ in order to “pinpoint” the entropy loss in $\sigma_t(X_1) \oplus X_2$. That is, as a result of the first two steps we “trade the noise-resistance off” for localising the entropy loss in $\sigma_t(X_1) \oplus X_2$.

In Step 3 we apply hypercontractivity to conclude that $\nu$ cannot have the full entropy, quod erat demonstrandum.

**Step 1**

If a distribution is noise-resistant, then its “projection” to a random $2/3$-fraction of bits is also distinguishable from the uniform – this will let us conclude that under the assumptions of Lemma 3, entropy loss must occur in a “typical” $(\sigma_t(X_1) \oplus X_2)_S$ for $|S| = \lfloor 2n/3 \rfloor$. Formally:

**Lemma 4.** If $\nu$ is a distribution on $\{0, 1\}^n$, such that $H(T_{\delta/4}(\nu)) \leq n - \delta$ for $\delta > 0$, then

$$E_{\nu} \left[ H \left( (\sigma_t(X_1) \oplus X_2)_S \right) \right] \leq \left\lceil \frac{2n}{3} \right\rceil - \delta + 2\log n - \Omega(n),$$

where $S$ is uniformly distributed over the subsets of $\{1, \ldots, n\}$ of size $\lfloor 2n/3 \rfloor$.

Note that the distributions of $\sigma_t(X_1) \oplus T_{\delta/4}(X_2)$ and of $T_{\delta/4}(\sigma_t(X_1) \oplus X_2)$ are the same. Therefore, the above lemma implies that

$$E_{\nu} \left[ H \left( (\sigma_t(X_1) \oplus X_2)_S \right) \right] \leq \left\lceil \frac{2n}{3} \right\rceil - (n - H \left( \sigma_t(X_1) \oplus T_{\delta/4}(X_2) \right)) + 2\log n - \Omega(n).$$

From the assumption of Lemma 3 it follows that

$$E_{\nu} \left[ \left\lceil \frac{2n}{3} \right\rceil - \delta + 2\log n - \Omega(n). \right)$$

**Proof of Lemma 4.** For any $x_0 \in \{0, 1\}^n$, $Y \sim T_{\delta/4}(x_0)$ can be sampled via the following procedure.\(^\text{13}\) Pick independently $W, Z \sim \mathcal{U}(\{0, 1\}^n)$; on those coordinates $j$ where $W_j = 1$, let $Y_j = x_0(j)$; otherwise, let $Y_j = Z_j$. The same method can be used to sample $Y \sim T_{\delta/4}(\nu)$:

$$Y = (W \land X) \lor (\neg W \land Z),$$

where $X \sim \nu$ and $W, Z \sim \mathcal{U}(\{0, 1\}^n)$ independently.

By the non-growth of entropy under conditioning (1),

$$n - \delta \geq H(Y) \geq H(Y|W) = H(X_W|W) + H(Z_W|W) = H(X_W|W) + E[n - |W|].$$

\(^\text{13}\)This intuitive and convenient way of looking at the noisy distribution is used in [Sam15] for very similar purpose – to analyse the noise-resistance of a Boolean function (cf. Remark 1.10 there).
Note that
\[ H(X_W|W) + E[n - |W|] = \sum_{k=0}^{n} \Pr(|W| = k) \cdot \left( E_{|S|=k} \left[ H(X_S|S) \right] + n - k \right) \]
and that \( E_{|S|=k} \left[ H(X_S|S) \right] + n - k \) is non-increasing in \( k \), as
\[ E_{|S|=k+1} \left[ H(X_S|S) \right] - E_{|S|=k} \left[ H(X_S|S) \right] \leq 1. \]
Accordingly,
\[ n - \delta \geq \left( E_{|S|=[2n/3]} \left[ H(X_S|S) \right] + n - \left\lceil \frac{2n}{3} \right\rceil \right) \cdot \left( 1 - \Pr(|W| > \left\lceil \frac{2n}{3} \right\rceil) \right) \]
\[ \geq \left( E_{|S|=[2n/3]} \left[ H(X_S|S) \right] + n - \left\lceil \frac{2n}{3} \right\rceil - n \cdot 2^{-\Omega(n)} \right), \]
where the last inequality follows from the Chernoff bound. \( \blacksquare \) \textbf{Lemma 4}

\textbf{Step 2}

Here we will use the chain decomposition and the non-growth of entropy under conditioning in order to “pinpoint” the entropy loss in \( \sigma_i(X_1) \oplus X_2 \).

Let us look at
\[ E_{|S|=[2n/3]} \left[ H(Y_S|S) \right] \]
for some \( Y \), taking values from \( \{0, 1\}^n \):
\[ E_{|S|=[2n/3]} \left[ H(Y_S|S) \right] = E_{\tau \in S_n} \left[ \sum_{i=1}^{\left\lceil \frac{2n}{3} \right\rceil} H(Y_{\tau(i)}|Y_{\tau(1)}, \ldots, Y_{\tau(i-1)}) \right] \]
\[ = \sum_{i=1}^{\left\lceil \frac{2n}{3} \right\rceil} E_{\tau \in S_n} \left[ H(Y_{\tau(i)}|Y_{\tau(1)}, \ldots, Y_{\tau(i-1)}) \right] \]
\[ = \sum_{i=0}^{\left\lceil \frac{2n}{3} \right\rceil - 1} E_{j \in \{1, \ldots, n\}} \sum_{S \subseteq \{1, \ldots, n\} \setminus \{j\}}^{\left\lceil \frac{2n}{3} \right\rceil - 1} \left[ H(Y_j|S, Y_S) \right] \]
\[ \geq \left\lceil \frac{2n}{3} \right\rceil \cdot E_{j \in \{1, \ldots, n\}} \sum_{S \subseteq \{1, \ldots, n\} \setminus \{j\}}^{\left\lceil \frac{2n}{3} \right\rceil - 1} \left[ H(Y_j|S, Y_S) \right], \]
where the inequality follows from the non-growth of entropy under conditioning (1).

Substituting \( Y \overset{\text{def}}{=} (\sigma_i(X_1) \oplus X_2) \) and combining the above with (5) (under \( i \in \{0, \ldots, n - 1\} \)), we get that for large enough \( n \),
\[ E_{i \in \{0, \ldots, n-1\}, j \in \{1, \ldots, n\}, S \subseteq \{1, \ldots, n\} \setminus \{j\}, |S|=[2n/3]} \left[ H_{(X_1,X_2) \sim \nu} \left( (\sigma_i(X_1) \oplus X_2)_j \right| S, (\sigma_i(X_1) \oplus X_2)_S \right) \right] \leq 1 - \frac{\delta}{n}, \]
From the non-growth of entropy under conditioning (1) it further follows that

\[
1 - \frac{\delta}{n} \geq \min_{i \in \{0, \ldots, n-1\} \atop j \in \{1, \ldots, n\} \atop \mathcal{S} \subseteq \{1, \ldots, n\} \setminus \{j\} \atop |\mathcal{S}| = \lceil 2n/3 \rceil} E_{i,j,\mathcal{S}} \left[ H \left( (\sigma_i(X_1) \oplus X_2)_j \mid S, (\sigma_i(X_1) \oplus X_2)_S \right) \right] \\
\geq \min_{i,j,S} E_{i,j,S} \left[ H \left( (\sigma_i(X_1) \oplus X_2)_j \mid S, (\sigma_i(X_1))_S, X_2(S) \right) \right] \\
= \min_{i,j,S} E_{i,j,S} \left[ H \left( X_1(\sigma_{-i}(j)) \oplus X_2(j) \mid S, (\sigma_i(X_1))_S, X_2(S) \right) \right]. \tag{6}
\]

For any \(i_0\), at most \(\lceil n/3 \rceil\) pairs of the form \((\sigma_{-i_0}(j), j)\) can be outside of \(S_1 \times S_2\) when \(|S_1| = |S_2| = \lceil 5n/6 \rceil\). Therefore, for uniformly-random \(S_1, S_2\) and \(S_2\) conditioning on \([S_1, S_2, X_1(S_1), X_2(S_2)]\) is at least as informative as conditioning on \([S, (\sigma_i(X_1))_S, X_2(S)]\), and so, from the non-growth of entropy under conditioning (1) it follows that the expectation (6) is at least

\[
\min_{i \in \{0, \ldots, n-1\} \atop j \in \{1, \ldots, n\} \atop \mathcal{S}_1 \subseteq \{1, \ldots, n\} \setminus \{j\} \atop |\mathcal{S}_1| = |\mathcal{S}_2| = \lceil 2n/3 \rceil} E_{i,j,\mathcal{S}_1,\mathcal{S}_2} \left[ H \left( X_1(j_1) \oplus X_2(j_2) \mid S_1, S_2, X_1(S_1), X_2(S_2) \right) \right] \\
= \min_{j_1 \in \{1, \ldots, n\} \setminus \mathcal{S}_1 \atop j_2 \in \{1, \ldots, n\} \setminus \mathcal{S}_2} E_{j_1,\mathcal{S}_1,j_2,\mathcal{S}_2} \left[ H \left( X_1(j_1) \oplus X_2(j_2) \mid S_1, S_2, X_1(S_1), X_2(S_2) \right) \right].
\]

So,

\[
\min_{S_1, S_2 \subseteq \{1, \ldots, n\} \atop |S_1| = |S_2| = \lceil n/6 \rceil} E_{S_1,S_2} \left[ H \left( X_1(j_1) \oplus X_2(j_2) \mid S_1, S_2, X_1(S_1), X_2(S_2) \right) \right] \leq 1 - \frac{\delta}{n},
\]

in particular, there exist subsets \(T_1, T_2 \subseteq \{1, \ldots, n\}\) of size \(\lceil n/6 \rceil\), such that

\[
\min_{j_1 \in I_1} \min_{j_2 \in T_2} E_{j_1,j_2} \left[ H \left( X_1(j_1) \oplus X_2(j_2) \mid X_1(T_1), X_2(T_2) \right) \right] \leq 1 - \frac{\delta}{n},
\]

where \(\overline{T_i}\) stands for \(\{1, \ldots, n\} \setminus T_i\).

Let \(I_1 \equiv T_1, I_2 \equiv n + T_2\), \(I \equiv I_1 \cup I_2\) and \(T \equiv \{1, \ldots, 2n\} \setminus I\). Then

\[
\frac{\delta}{n} \leq \min_{j_1 \in I_1} \min_{j_2 \in I_2} E_{j_1,j_2} \left[ 1 - H \left( Y_{j_1} \oplus Y_{j_2} \mid Y_T \right) \right] \leq 2 \cdot \min_{j_1 \neq j_2 \in I} E_{j_1,j_2} \left[ 1 - H \left( Y_{j_1} \oplus Y_{j_2} \mid Y_T \right) \right], \tag{7}
\]

where \(|I| \geq \lceil n/3 \rceil - 2\).
Step 3

Finally, we apply hypercontractivity to argue that if (7) holds then $\nu$ cannot have the full entropy. Formally:

**Lemma 5.** If $\rho$ is a distribution on $\{0,1\}^n$ for large enough $n$, then

$$
\mathbb{E}_{j_1 \neq j_2 \in \{1,\ldots,n\}}\left[1 - H_{Y \sim \nu} (X_{j_1} \oplus X_{j_2})\right] \leq \frac{45}{n^2} \left(n - H_{\min}(\rho)\right)^2.
$$

Using it, we can write (assuming large enough $n$):

$$
\mathbb{E}_{W \sim \nu}\left[\left(|I| - H_{Y \sim \nu} (Y_I | Y_T = W_T)\right)^2\right] \\
\geq \frac{|I|^2}{45} \cdot \mathbb{E}_{W \sim \nu}\left[\mathbb{E}_{j_1 \neq j_2 \in I} \left[1 - H_{Y \sim \nu} (Y_{j_1} \oplus Y_{j_2} | Y_T = W_T)\right]\right] \\
\geq \frac{n^2}{406} \cdot \mathbb{E}_{j_1 \neq j_2 \in I} \left[1 - H_{Y \sim \nu} (Y_{j_1} \oplus Y_{j_2} | Y_T)\right] \\
\geq \frac{\delta n}{812},
$$

where the first inequality is an application of Lemma 5 with respect to $Y_I$ taking values from $\{0,1\}^{|I|}$, conditioned on the given value of $Y_T$, the second one uses the fact that $|I| \geq n/3 - 2$; the last one follows from (7) and the fact that $H_{\min}$ is never greater than $H$. Using Claim 1 with $A$ being the support of $Y_T$, $B$ being the support of $Y_I$ and $\Delta = 2 \log n + 2$, we get

$$
\frac{1}{4n^2} \geq \mathbb{P}_{W \sim \nu}\left[H_{Y \sim \nu} (Y_I | Y_T = W_T) \leq H_{\min}(\nu) - (2n - |I|) - 2 \log n - 2\right]
$$

$$
= \mathbb{P}_{W \sim \nu}\left[|I| - H_{Y \sim \nu} (Y_I | Y_T = W_T) \geq 2n - H_{\min}(\nu) + 2 \log n + 2\right],
$$

and so,

$$
\frac{\delta n}{812} \leq \mathbb{E}_{W \sim \nu}\left[\left(|I| - H_{Y \sim \nu} (Y_I | Y_T = W_T)\right)^2\right] \leq \left(2n - H_{\min}(\nu) + 2 \log n + 2\right)^2 + 1.
$$

Lemma 3 follows.

**Proof of Lemma 5.** For $j_1 \neq j_2 \in \{1,\ldots,n\}$, let $\Delta_{j_1,j_2} \overset{\text{def}}{=} |\mathbb{P}_\rho [X_{j_1} = X_{j_2}] - 1/2|$. Then

$$
\mathbb{E}_{j_1 \neq j_2} \left[1 - H_{\rho} (X_{j_1} \oplus X_{j_2})\right] \leq \frac{\mathbb{P}_{j_1 \neq j_2} \left[\Delta_{j_1,j_2} > 1/4\right]}{6 \cdot \mathbb{E}_{j_1 \neq j_2} \left[\Delta_{j_1,j_2}^2\right]} < 22 \mathbb{E}_{j_1 \neq j_2} \left[\Delta_{j_1,j_2}^2\right],
$$

where the first inequality follows from the Taylor expansion of the binary entropy function (2):

$$
1 - H_{\rho} (X_{j_1} \oplus X_{j_2}) = \frac{1}{2 \ln 2} \cdot \sum_{i=1}^{\infty} \frac{(2\Delta_{j_1,j_2})^{2i}}{i(2i - 1)} \leq \frac{2\Delta_{j_1,j_2}^2}{\ln 2} \cdot \sum_{i=0}^{\infty} (2\Delta_{j_1,j_2})^{2i} \\
= \frac{2\Delta_{j_1,j_2}^2}{\ln 2 \cdot (1 - 4\Delta_{j_1,j_2}^2)} \leq \frac{2\Delta_{j_1,j_2}^2}{\ln 2 \cdot (1 - 2\Delta_{j_1,j_2})},
$$

18
where the right-hand side is at most $6 \cdot \Delta^2_{j_1,j_2}$ when $\Delta_{j_1,j_2} \leq 1/4$, and therefore,

$$1 - H_\rho(X_{j_1} \oplus X_{j_2}) \leq \begin{cases} 6 \cdot \Delta^2_{j_1,j_2} & \text{if } \Delta_{j_1,j_2} \leq 1/4; \\ 1 & \text{always.} \end{cases}$$

On the other hand,

$$\Delta_{j_1,j_2} = \frac{1}{2} \cdot 2 \left| \mathbb{P}_\rho[X_{j_1} = X_{j_2}] - 1 \right| = \frac{1}{2} \cdot \left| \mathbb{E}_\rho[\chi_{\{j_1,j_2\}}(x)] \right|

= \frac{1}{2} \cdot \left| \sum_x \rho(x) \chi_{\{j_1,j_2\}}(x) \right| = 2^{n-1} \cdot |\hat{\rho}(\{j_1,j_2\})|,$$

and from (8),

$$\mathbb{E}_{j_1 \neq j_2} \left[ 1 - H_\rho(X_{j_1} \oplus X_{j_2}) \right] < 22 \cdot \mathbb{E}_{j_1 \neq j_2} \left[ \Delta^2_{j_1,j_2} \right] < 6 \cdot \frac{2^n}{(2^2)} \cdot \sum_{|s|=2} \hat{\rho}(s)^2.$$

From Claim 3 with $t = 2$,

$$\sum_{|s|=2} \hat{\rho}(s)^2 \leq \|\rho\|_1^2 \cdot \left( e \cdot \ln \frac{\|\rho\|_\infty}{\|\rho\|_1} \right)^2 = 2^{-2n} \cdot \left( \frac{e}{\log e} \right)^2 \cdot \left( n - H_{\min}(\rho) \right)^2,$$

and the result follows. $\blacksquare$

## 7 Discussion

We have shown that the partial function $\text{Shape}$ is easy for $Q^{1, \text{ent}}$ but hard for $R$, which may be viewed as improving earlier understanding of when quantum communication can outperform classical communication. Nevertheless, prior to this work there have been a number of results that have emphasised the advantage of quantum communication: [Raz99, BCWdW01, BYJK04, GKK+08, Gav08, KR11] and many others.

On the other hand, current knowledge of when classical communication can “replace” quantum communication – i.e., when the advantage of using quantum communication can be at most (quasi-) polynomial in terms of complexity – is surprisingly limited (cf. [Aar04, GRdW08]). We do not want to speculate now whether or not it is possible, say, to find a partial function, easy for $Q^{1}$ but hard for $R$ – finding that out is, obviously, an important open problem; however, our intuition strongly suggests that no total function can have that property. Proving or refuting it would be a breakthrough.

On a different note, several questions regarding the complexity of $\text{Shape}$ in other models of communication are still open. Answering any of the following will necessarily lead to a yet-unknown (as of now) separation of communication complexity classes.

- What is the $Q^{1}$-complexity of $\text{Shape}$? If it has an efficient protocol, that would strengthen qualitatively the result of this work. If it is hard, that would give the first superpolynomial separation between $Q^{1, \text{ent}}$ and $Q^{1}$ via a partial function (a relational separation is known).

14 $Q^{1}$ is the model of (“unentangled”)
quantum simultaneous message passing; closely related to it is $Q^{1, \text{pub}}$ – the model of quantum simultaneous message passing with shared randomness.
What is the $Q^{∥,\text{pub}}$-complexity\textsuperscript{14} of Shape? The consequences would be similar.

What is the $R^{∥,\text{ent}}$-complexity\textsuperscript{15} of Shape? The consequences would be similar, and not even a relational separation is currently known between $Q^{∥,\text{ent}}$ and $R^{∥,\text{ent}}$.

Acknowledgements

I am very grateful to Oded Regev, Ronald de Wolf and anonymous reviewers for many helpful comments.

References

\textsuperscript{14} $Q^{∥,\text{pub}}$ is the model of quantum simultaneous message passing with public randomness.

\textsuperscript{15} $R^{∥,\text{ent}}$ is the model of classical simultaneous message passing with entanglement.


