The space complexity of recognizing well-parenthesized expressions in the streaming model: the Index function revisited*

Rahul Jain† Ashwin Nayak‡

March 6, 2014

Abstract

We show an $\Omega(\sqrt{n}/T)$ lower bound for the space required by any unidirectional constant-error randomized T-pass streaming algorithm that recognizes whether an expression over two types of parenthesis is well-parenthesized. This proves a conjecture due to Magniez, Mathieu, and Nayak (2009) and rigorously establishes that bidirectional streams are exponentially more efficient in space usage as compared with unidirectional ones. We obtain the lower bound by analyzing the information that is necessarily revealed by the players about their respective inputs in a two-party communication protocol for a variant of the Index function, namely Augmented Index. We show that in any communication protocol that computes this function correctly with constant error on the uniform distribution (a "hard" distribution), either Alice reveals $\Omega(n)$ information about her n-bit input, or Bob reveals $\Omega(1)$ information about his $(\log n)$-bit input, even when the inputs are drawn from an "easy" distribution, the uniform distribution over inputs which evaluate to 0. The information cost trade-off is obtained by a novel application of the conceptually simple and familiar ideas such as *average encoding* and the *cut-and-paste property* of randomized protocols.

Motivated by recent examples of exponential savings in space by streaming *quantum* algorithms, we also study quantum protocols for Augmented Index. Defining an appropriate notion of information cost for quantum protocols involves a delicate balancing act between its applicability and the ease with which we can analyze it. We define a notion of quantum information cost which reflects some of the non-intuitive properties of quantum information. We show that in quantum protocols that compute the Augmented Index function correctly with constant error on the uniform distribution, either Alice reveals $\Omega(n/t)$ information about her n-bit input, or Bob reveals $\Omega(1/t)$ information about his $(\log n)$-bit input, where t is the number of messages in the protocol, even when the inputs are drawn from the abovementioned easy distribution. While

*The results on quantum communication in this article were presented at the 15th Workshop on Quantum Information Processing, QIP 2012, Dec., 2011.

†Centre for Quantum Technologies and Department of Computer Science, S15 #04-01, 3 Science Drive 2, National University of Singapore, Singapore 117543. Email: rahul@comp.nus.edu.sg. Work done in part while visiting Institute for Quantum Computing, University of Waterloo. This work is supported by the Singapore Ministry of Education Tier 3 Grant and the Core Grants of the Center for Quantum Technologies, Singapore.

‡Department of Combinatorics and Optimization, and Institute for Quantum Computing, University of Waterloo, 200 University Ave. W., Waterloo, ON, N2L 3G1, Canada. Email: ashwin.nayak@uwaterloo.ca. Work done in part at Perimeter Institute for Theoretical Physics, and while visiting Center for Quantum Technologies, National University of Singapore. Research supported in part by NSERC Canada, CIFAR, an ERA (Ontario), QuantumWorks, MITACS, and ARO (USA). Research at Perimeter Institute is supported in part by the Government of Canada through Industry Canada and by the Province of Ontario through MRI.

this trade-off demonstrates the strength of our proof techniques, it does not lead to a space lower bound for checking parentheses. We leave such an implication for quantum streaming algorithms as an intriguing open question.

Keywords: streaming algorithm, space complexity, Dyck language, communication complexity, information cost, Augmented Index, quantum information theory, quantum communication

1 Introduction

Streaming algorithms [39] are designed to process massive input data, which cannot fit entirely in computer memory. Random access to such input is prohibitive, so ideally we would like to process it with a single sequential scan. Furthermore, during the computation, the algorithms are compelled to use space that is much smaller than the length of the input. Formally, streaming algorithms access the input sequentially, one symbol at a time, a small number of times (called passes), while attempting to solve some information processing task using as little space (and time) as possible.

One-pass streaming algorithms that use constant space and time recognize precisely the set of regular languages. It is thus natural to ask what the complexity of languages higher up in the Chomsky hierarchy is in the streaming model. In this work, we focus on a concrete such problem, that of checking whether an expression with different types of parenthesis is well-formed. The problem is formalized through the language DYCK(2), which consists of all well-parenthesized expressions over two types of parenthesis, denoted below by $a, \overline{a}$ and $b, \overline{b}$, with the bar indicating a closing parenthesis. Formally, DYCK(2) is the language over alphabet $\Sigma = \left\{a, \overline{a}, b, \overline{b}\right\}$ defined recursively as

$$\text{DYCK}(2) \quad = \quad \epsilon + \left(a \cdot \text{DYCK}(2) \cdot \overline{a} + b \cdot \text{DYCK}(2) \cdot \overline{b}\right) \cdot \text{DYCK}(2) \ ,$$

where ϵ is the empty string, '$\cdot$' indicates concatenation of strings (or subsets thereof) and '+' denotes set union. This deceptively simple language is in a certain precise sense complete for the class of context-free languages [14], and is implicit in a myriad of information processing tasks.

There is a straightforward algorithm that recognizes DYCK(2) with logarithmic space, as we may run through all possible levels of nesting, and check parentheses at the same level. While this scheme is highly space-efficient, it may make $\Omega(n)$ passes over the input in the worst case, on instances of length n. It is not obvious if we can translate this scheme to a streaming algorithm with a small number of passes over the input. By appealing to the communication complexity of the equality function, we can deduce that any *deterministic* streaming algorithm for DYCK(2) that makes T passes over the input requires space $\Omega(n/T)$ on instances of length n. Therefore, any streaming algorithm with smaller space complexity, if one exists, would necessarily be randomized. One such algorithm is suggested by a small-space algorithm for the word problem in the free group with 2 generators. This is a relaxation of DYCK(2) in which local simplifications $\bar{p}p = \epsilon$ are allowed in addition to $p\bar{p} = \epsilon$ for every type of parenthesis $(p, \bar{p})$. There is a logarithmic space (randomized) algorithm for solving the word problem [36] that can easily be massaged into a one-pass streaming algorithm with polylogarithmic space. Again, this algorithm does not extend to DYCK(2).

We rigorously establish the impossibility of recognizing DYCK(2) with logarithmic space with a small number of passes in the streaming model, even with randomized algorithms.

Theorem 1.1. *For any $T \geq 1$, any unidirectional randomized T-pass streaming algorithm that recognizes length n instances of* DYCK(2) *with a constant probability of error uses space $\Omega(\sqrt{n}/T)$.*

A more precise statement of this theorem is presented as Corollary 3.3 later in this article.

DYCK(2) was first studied in the context of the streaming model by Magniez, Mathieu, and Nayak [37]. They were motivated by its practical relevance, e.g., its relationship to the processing of large XML files, and by the connection between formal language theory and complexity

in the context of processing massive data. They overcome the apparent difficulties described above and present sublinear space randomized streaming algorithms for DYCK(2). The first makes *one* pass over the input, recognizes well-parenthesized expressions with space $O(\sqrt{n \log n})$ bits, and has polynomially small probability of error. Moreover, they prove that this one-pass algorithm is optimal. They establish that any one-pass randomized algorithm that makes error at most $1/n \log n$ uses space $\Omega(\sqrt{n \log n})$. Theorem 1.1 establishes a similar result for *multi-pass* streaming algorithms. The bound for one-pass algorithms given by Theorem 1.1 is a factor of $\sqrt{\log n}$ better than the one in Ref. [37] for constant error probability, but falls short of optimal (by the same factor) for polynomially small error.

In the standard model for streaming algorithms, access to the input symbols is provided in the *same fixed order in every pass over the input*. This reflects a constraint of the infrastructure available to us in practice. Theorem 1.1 applies to such *unidirectional* algorithms. Perhaps surprisingly, Magniez *et al.* showed that the demand on space shrinks drastically when algorithms for DYCK(2) are allowed another pass over the input in the *reverse* direction. They presented a second algorithm that makes two passes in opposite directions over the input, uses only $O(\log^2 n)$ space, and has polynomially small probability of error. A question that naturally arose is whether this is an artefact of the algorithm, or if we could achieve similar reduction in space usage by making multiple passes in the same direction. Magniez *et al.* conjecture that a bound similar to that for the one-pass algorithms hold for multi-pass streaming algorithms if all passes are made in the same direction. Theorem 1.1 proves this conjecture and establishes the first natural example for which unidirectional multi-pass streaming algorithms are much less powerful than bidirectional ones. More importantly, existing computing infrastructure only supports unidirectional streams, and this result confirms that we cannot reproduce the performance of the bidirectional algorithm within it.

Theorem 1.1 is a consequence of a lower bound that we establish for the "information cost" of two-party communication protocols for a variant of the INDEX problem. In the INDEX problem, one party, Alice, is given an n-bit string x, and the other party, Bob, is given an integer $k \in [n]$. Their goal is to determine the bit x_k by communicating with each other. In the variant we study, the player holding the index also receives a portion of the other party's input. More formally, Alice holds an n-bit string x, and Bob, holds an integer $k \in [n]$, the prefix $x[1, k-1]$ of x, and a bit $b \in \{0, 1\}$. The goal is to compute the function $f_n(x, (k, x[1, k-1], b)) = x_k \oplus b$, i.e., to determine whether $b = x_k$ or not. This problem was studied in the one-way communication model, with communication from Alice to Bob, as "serial encoding" [2, 40]. Lower bounds on its quantum communication complexity were derived and used to establish exponential lower bounds on the size of one-way quantum finite automata. In later works, the problem was studied as "Augmented Index"; the linear lower bound was re-derived for classical communication, and used to establish lower bounds for streaming and sketching (see, e.g., [26, 16]). The problem, called "the Mountain problem" by Magniez, Mathieu, and Nayak [37], was central to the proof of optimality of the one-pass streaming algorithm for DYCK(2). We elaborate on this later in this section.

Informally speaking, we show that in any communication protocol that computes the AUGMENTED INDEX function f_n with constant error on the uniform distribution μ (a "hard distribution"), either Alice reveals $\Omega(n)$ information about her n-bit input x, or Bob reveals $\Omega(1)$ information about his $(\log n)$-bit input k, even when the inputs are drawn from an "easy distribution" (μ_0, the uniform distribution over $f_n^{-1}(0)$). We formally define the notion of information cost $(\mathrm{IC}^{\mathsf{A}}_{\lambda}(\Pi), \mathrm{IC}^{\mathsf{B}}_{\lambda}(\Pi))$ for a protocol Π for the two players Alice (A) and Bob (B) with respect to the distribution λ in Section 2.3, and show:

Theorem 1.2. *In any two-party randomized communication protocol* Π *for the* AUGMENTED INDEX *function* f_n *that makes constant error at most* $\varepsilon \in [0, 1/4)$ *on the uniform distribution* μ *over inputs, either* $\mathrm{IC}^{\mathsf{A}}_{\mu_0}(\Pi) \in \Omega(n)$ *or* $\mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi) \in \Omega(1)$.

A more precise statement of this theorem is presented as Theorem 2.6 later in this article. We point out that the theorem is optimal as there is a one-message deterministic protocol for AUGMENTED INDEX with communication n.

The connection between streaming algorithms using "small" space to two-party protocols for AUGMENTED INDEX with "small" information cost was presented by Magniez *et al.* for one-pass algorithms. However, it generalizes in a straightforward manner to multi-pass algorithms. For completeness, this reduction is described in full in Section 3, for multi-pass algorithms. The reduction consists of three steps, following the information cost approach. (See, for example, Refs. [13, 45, 5, 25, 23] for earlier applications of this approach.) First, a streaming algorithm for DYCK(2) that uses space s is mapped to a multi-party communication protocol in which the messages are each of the same length s. Second, a two-party communication protocol for AUGMENTED INDEX with "small" information cost with respect to μ_0 is derived using a "direct sum" argument. Finally, a lower bound for the aforementioned information cost is proven. Magniez *et al.* proved a lower bound for the information cost of a *two-message* protocol that resulted from a one-pass streaming algorithm. Our main contribution, Theorem 1.2, lies in this final step. It applies to protocols with an arbitrary number of messages, and is the first general lower bound on information cost for AUGMENTED INDEX.

A notion of information cost for INDEX was studied previously by Jain, Radhakrishnan, and Sen [24] in the context of privacy in communication (see also earlier work due to Klauck [28]). This notion differs from the one we study in two crucial respects. First, it is defined in terms of the hard distribution for the problem (uniform over all inputs). Second, the hard distribution is a product distribution. The techniques they develop seem not to be directly relevant to the problem at hand, as we deal with an easy and non-product distribution.

We devise a new method for analyzing the information cost of f_n to arrive at Theorem 1.2. The proof we present shows how conceptually simple and familiar ideas such as *average encoding* and the *cut-and-paste property* of randomized protocols may be brought to bear on AUGMENTED INDEX to derive the optimal (up to constant factors) information cost trade-off. The intuition behind the lower bound is as follows. Assume, for simplicity, that the protocol transcript contains the output. Starting from an input pair on which the function evaluates to 0, if the information cost of any one party is "low" and we carefully change her input, the transcript does not change "much". We show that even when we simultaneously change the inputs with both parties, resulting in a 1-input of the function, the perturbation to the transcript state is also correspondingly "small". This implies that the two information costs cannot be "small" simultaneously.

We point out that the trade-off established by Magniez, Mathieu, and Nayak [37] for *two-message* protocols that *start with Alice*, and make polynomially small error, is stronger. They show that either Alice reveals $\Omega(n)$ information about x, or Bob reveals $\Omega(\log n)$ information about k in such protocols. This cannot be reproduced without a further refinement of our techniques. Indeed, Theorem 1.2 also applies to two-message protocols in which *Bob* starts. Such protocols match the trade-off given in the theorem: for every $l \in \{1, 2, \ldots, \lfloor \log_2 n \rfloor\}$, there is a deterministic protocol for f_n in which Bob sends l bits of k, and Alice responds with $n/2^l$ bits.

In independent work, concurrent with ours, Chakrabarti, Cormode, Kondapally, and McGregor [11] derive a similar information cost trade-off for f_n. Their motivation is identical to ours—to study the space required by unidirectional multi-pass streaming algorithms for DYCK(2), and they present a similar space lower bound for such algorithms. While some of the basic tools from information theory at the heart of their proof (e.g., the Chain Rule for mutual information and the Pinsker Inequality) are equivalent to ours, they take a different route to these tools. The first version of our article [22] and that of Chakrabarti *et al.* [10] contained trade-offs that were weaker, albeit in different respects. After learning about each other's work, both groups strengthened our respective proofs to achieve qualitatively the same result. Subsequently, Chakrabarti and Kondapally [12] extended the result to show that either Bob reveals $\Omega(b)$ information about his input k, or Alice reveals $n/2^{\mathrm{O}(b)}$ information about her input x, i.e., either $\mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi) \in \Omega(b)$ or $\mathrm{IC}^{\mathsf{A}}_{\mu_0}(\Pi) \in n/2^{\mathrm{O}(b)}$. This matches the information cost of the two-message protocol described above up to constant factors.

The promise of fast processing with limited memory held by streaming algorithms make them especially attractive in the context of quantum computation. The absence of prototypes

with a large enough number of qubits and long coherence times inevitably leads us to such algorithms. This has fueled the study of quantum finite automata and also later works on quantum streaming algorithms [34, 21, 8]. Several of these works show how quantum effects lead to an exponential savings in space over their classical counterparts, albeit for specially crafted problems. It is thus natural to ask how much more efficient such quantum algorithms could be, for a well-studied and important problem such as DYCK(2). Motivated by this, we also study quantum protocols for AUGMENTED INDEX. We define appropriate notions of quantum information cost $(\mathrm{QIC}^{\mathsf{A}}_{\lambda}(\Pi), \mathrm{QIC}^{\mathsf{B}}_{\lambda}(\Pi))$ for distributions λ with a limited form of dependence in Section 4.2, and then arrive at the following trade-off.

Theorem 1.3. *In any two-party quantum communication protocol* Π *(with read-only behaviour on inputs and no intermediate measurements) for the* AUGMENTED INDEX *function* f_n *that has* t *message exchanges and makes constant error at most* $\varepsilon \in [0, 1/4)$ *on the uniform distribution* μ *over inputs, either* $\mathrm{QIC}^{\mathsf{A}}_{\mu_0}(\Pi) \in \Omega(n/t)$ *or* $\mathrm{QIC}^{\mathsf{B}}_{\mu_0}(\Pi) \in \Omega(1/t)$.

Quantum protocols have the ability to compute without revealing much information [20, 18]. It is thus hardly a surprise that the quantum information cost trade-off involves a number of subtleties. For instance, it is not obvious how we may quantify information cost in the absence of the notion of a message transcript, or how we discount information leakage due to the non-product nature of the input distribution. These issues are discussed in detail in Section 4.2. Nonetheless, we show how the ideas behind Theorem 1.2 also shed light on quantum communication. The intuition from the classical case comes with its own complications, such as the absence of an analogue of the Cut-and-Paste Lemma. We circumvent the Cut-and-Paste property by appealing to the "Local Transition Theorem" and adapting a hybrid argument due to Jain, Radhakrishnan, and Sen [23]. We apply these on a message-by-message basis, which leads to the dependence of the trade-off on the number of messages in the protocol. We are not aware of quantum protocols that beat the classical information bounds. However the dependence of the trade-off in Theorem 1.3 on the number of messages t may be inherent, as is the case with Set Disjointness [23].

Theorem 1.3 demonstrates the versatility of our proof techniques. The techniques due to Magniez *et al.* [37] and Chakrabarti *et al.* [11] for showing information cost trade-off in classical protocols do not seem to generalize to quantum protocols. They analyze the input distribution conditioned on the message transcript, a notion for which no suitable quantum analogue is known. Theorem 1.3, however, does not immediately lead to a lower bound on the space required by quantum streaming algorithms for DYCK(2). The main hurdle here is that the connection between streaming algorithms and communication protocols for AUGMENTED INDEX with low information cost does not extend to the quantum case. This appears to be due to the stronger notion of information cost that we adopt. (The stronger notion appears to be necessary for our proof technique.) It is possible that a version of Theorem 1.3 hold with an alternative definition of information cost that is more relevant to quantum streaming algorithms. We leave this for future investigation.

Communication problems involving the INDEX and AUGMENTED INDEX functions capture a number of phenomena in the theory of computing, both classical and quantum, in addition to playing a fundamental role in the area of communication complexity [32]. For instance, they have been used to analyze data structures [38], the size of finite automata [3] and formulae [29], the length of locally decodable codes [27], learnability of states [31, 1], and sketching complexity [4]. Recently, phenomena in quantum information have been discovered via the INDEX function problem, e.g., information causality [44], a connection between non-locality and the uncertainty principle [43] and quantum ignorance [47]. We believe that the more nuanced properties of the AUGMENTED INDEX function such as the one we establish here are of fundamental importance, and are likely to find application in other contexts as well.

Acknowledgments

We thank Frédéric Magniez and Christian Konrad for their comments on an earlier version of this article. A.N. thanks Frédéric Magniez also for several helpful discussions preceding this work.

We thank the authors of Ref. [11] for sending us their initial manuscript when we first publicized an earlier version of the article. The (classical) results in our respective articles were originally weaker in incomparable ways, and the exchange inspired both groups to refine our analyses to obtain the current classical information cost trade-off results.

We are grateful to the anonymous referees for their help in improving the presentation.

2 Classical information cost of Augmented Index

In this section we present the first result of this article. We summarize the notational conventions we follow and the background from classical information theory that we assume in Section 2.1. We do the same for two-party communication complexity and information cost in Section 2.2. Then we develop the lower bound for classical protocols for Augmented Index in Section 2.3.

2.1 Information theory basics

We reserve small case letters like x, k, m for bit-strings or integers, and capital letters like X, K, M for random variables over the corresponding sample spaces. We use the same symbol for a random variable and its distribution. As is standard, given jointly distributed random variables AB over a product sample space, A represents the marginal distribution over the first component. We sometimes use $A|b$ as shorthand for the conditional distribution $A|(B = b)$ when the second random variable B is clear from the context. For a string $x \in \{0,1\}^n$, and integers $i, j \in [n]$, where $[n] = \{1, 2, \ldots, n\}$, we let $x[i, j]$ denote the substring of consecutive bits $x_i \cdots x_j$. If $j < i$, the expression denotes the empty string. This notation extends to random variables over $\{0,1\}^n$ in the obvious manner. When a sample z is drawn from distribution Z, we denote it as $z \leftarrow Z$.

The ℓ_1 distance $\|A - B\|$ between two random variables A, B over the same finite sample space $\mathcal{S}$ is given by

$$\|A - B\| \quad = \quad \sum_{i \in \mathcal{S}} |A(i) - B(i)| \ ,$$

and takes values in the interval $[0, 2]$. (Recall that as per our notational convention $A(i), B(i)$ denote the probabilities assigned to $i \in \mathcal{S}$ by A, B, respectively.) The Hellinger distance $\mathfrak{h}(A\ ,\ B)$ between the random variables is defined as

$$\mathfrak{h}(A\ ,\ B) \quad = \quad \left[\frac{1}{2} \sum_{i \in \mathcal{S}} \left(\sqrt{A(i)} - \sqrt{B(i)}\right)^2\right]^{1/2} \ .$$

Hellinger distance is a metric, and is related to ℓ_1 distance in the following manner. (See Section 3.2 in [33] for a proof.)

Proposition 2.1. *Let P, Q be distributions over the same sample space. Then*

$$\mathfrak{h}(P\ ,\ Q)^2 \quad \leq \quad \frac{1}{2} \|P - Q\| \quad \leq \quad \sqrt{2}\, \mathfrak{h}(P\ ,\ Q) \ .$$

The square of the Hellinger distance satisfies the following property, called *joint convexity*. It may be verified by a straightforward application of the Cauchy-Schwarz inequality.

Proposition 2.2. *Let P_i, Q_i be distributions over the same sample space for each $i \in [n]$, and let (α_i) be a probability distribution over $[n]$. Let $P = \sum_{i=1}^n \alpha_i P_i$, and $Q = \sum_{i=1}^n \alpha_i Q_i$. Then*

$$\mathfrak{h}(P\ ,\ Q)^2 \quad \leq \quad \sum_{i=1}^n \alpha_i\, \mathfrak{h}(P_i\ ,\ Q_i)^2 \ .$$

Proof: By the Cauchy-Schwarz Inequality, for each $j \in \mathcal{S}$,

$$\begin{aligned}\sqrt{P(j)\,Q(j)} &= \left[\left(\sum_{i\in[n]} \alpha_i\, P_i(j)\right)\left(\sum_{i'\in[n]} \alpha_{i'}\, Q_{i'}(j)\right)\right]^{1/2} \\ &\geq \sum_{i\in[n]} \sqrt{\alpha_i\, P_i(j)}\sqrt{\alpha_i\, Q_i(j)}\ .\end{aligned}$$

So we have

$$\begin{aligned}\mathfrak{h}(P\ ,\ Q)^2 &= \frac{1}{2}\sum_{j\in\mathcal{S}}\left(P(j)+Q(j)-2\sqrt{P(j)\,Q(j)}\right) \\ &\leq \frac{1}{2}\sum_{j\in\mathcal{S}}\sum_{i\in[n]}\alpha_i\left(P_i(j)+Q_i(j)-2\sqrt{P_i(j)\,Q_i(j)}\right) \\ &= \sum_{i=1}^{n}\alpha_i\,\mathfrak{h}(P_i\ ,\ Q_i)^2\ .\end{aligned}$$

■

We rely on a number of standard results from information theory in this work. For a comprehensive introduction to the subject, we refer the reader to a text such as [15].

We use $\mathrm{H}(X)$ to denote the Shannon entropy of the random variable X, $\mathrm{I}(X:Y)$ to denote the mutual information between two random variables X, Y, and $\mathrm{I}(X:Y|Z)$ to denote the conditional mutual information of X, Y with respect to a jointly distributed random variable Z. We also use $\mathrm{H}(p)$ to denote the Binary entropy function when $p \in [0,1]$.

The chain rule for mutual information, Theorem 2.5.2 in [15], states:

Proposition 2.3 (Chain Rule)**.** *Let ABC be jointly distributed random variables. Then*

$$\mathrm{I}(AB:C) \quad = \quad \mathrm{I}(A:C)+\mathrm{I}(B:C\,|\,A)\ .$$

This implies that for jointly distributed random variables $A_1\cdots A_nC$,

$$\mathrm{I}(A_1\cdots A_n:C) \quad = \quad \mathrm{I}(A_1:C)+\mathrm{I}(A_2:C\,|\,A_1)+\cdots+\mathrm{I}(A_n:C\,|\,A_1\cdots A_{n-1})\ .$$

The Average encoding theorem [30, 23] is a quantitative version of the intuition that two random variables that are only weakly correlated are nearly independent. Stated differently, the conditional distribution of one given the other is close to its marginal distribution, if their mutual information is sufficiently small.

Proposition 2.4 (Average encoding theorem [30, 23])**.** *Let AB be jointly distributed random variables. Then,*

$$\mathbb{E}_{b\leftarrow B}\ \mathfrak{h}(A|b\ ,\ A)^2 \quad \leq \quad \kappa\,\mathrm{I}(A:B)\ ,$$

where κ is the constant $\frac{\ln 2}{2}$.

2.2 Communication protocols and information cost

In the two-party communication model [48] for computing Boolean functions, parties Alice and Bob receive inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, respectively, for some sets $\mathcal{X}, \mathcal{Y}$. They may share a random bit string R, that is independent of the inputs x, y. The bits of R are called *public* coins, as they are known to both parties. Alice (or Bob) may use an additional random string R_{A} (R_{B}, respectively), that is not known to the other party. These strings $R_{\mathsf{A}}, R_{\mathsf{B}}$ are called *private* coins.

The goal of the two parties is to compute a bi-variate Boolean function $f : \mathcal{X}\times\mathcal{Y} \to \{0,1\}$, by communicating with each other. The communication occurs in the form of $t \geq 0$ messages,

starting with one party, and then alternating with the other. In each of the t steps, the party sending it computes the message as a function of the input, the public and private random coins she or he has, and the messages received so far. After all t messages have been sent, the recipient of the last message produces the output of the protocol. The output is computed in a manner analogous to the messages, from the party's input, random coins, and all the messages received.

The pattern of communication is specified by a *protocol* Π, which lists the type, number, and distribution of the coins used by each party, the number of messages, the party that starts the protocol, and the functions used by the parties to generate the messages and the output. The sequence of t messages produced during a run of the protocol Π on a pair of inputs x, y together constitute the *transcript*. This is in general a random variable due to the use of random coins. We denote the random variable corresponding to the output by $\Pi(x, y)$. We point out that the transcript need not include the output of the protocol.

The probability of correctness (or *success*) of a protocol on input x, y is $\Pr[\Pi(x, y) = f(x, y)]$. We consider inputs drawn from a joint distribution XY, in which case the success probability is $\Pr[\Pi(X, Y) = f(X, Y)]$. The probability of the complementary event is called the *error* of the protocol on the distribution XY.

We refer the reader to the text [32] for equivalent formulations of communication protocols, and a thorough introduction to the models of two-party classical communication.

Protocols that use only public coins are called public-coin protocols and those that use only private coins are called private-coin protocols. The availability of public randomness obviates the need for private randomness in typical settings. Conversely, private randomness can often simulate public coins with a slight increase in communication [41]. In the context of information cost, however, access to the private randomness used by one party may result in more information being revealed to the other. To the best of our knowledge, there is no general recipe for replacing private with public randomness while preserving information cost. (For recent progress on this question, see Ref. [9].) In the reductions between protocols we encounter in this article, regardless of the nature of randomness used in the original protocol, we end up with a protocol with both types of randomness. We therefore study protocols of this type.

We use the following Cut-and-Paste property of private-coin communication protocols. (For a proof, see Lemma 6.3 in Ref. [5].)

Proposition 2.5 (Cut-and-Paste [5])**.** *Let* Π *be a two-party private-coin communication protocol. Let* $M(x, y)$ *denote the random variable representing the message transcript in* Π *when the first party has input* x *and the second party has input* y*. Then for all pairs of inputs* (x, y) *and* (u, v)*,*

$$\mathfrak{h}(M(x, y)\ ,\ M(u, v)) \quad = \quad \mathfrak{h}(M(x, v)\ ,\ M(u, y)) \quad .$$

We consider the information revealed during a communication protocol and focus on a notion known as "internal information" in the literature. Although this notion is implicit in earlier work [5], it was named so by Barak, Braverman, Chen, and Rao [6]. We emphasize that there is no canonical measure of information cost, and the choice of definition is often driven by a motivating application. A different definition of information cost would suffice for our application to streaming algorithms, and would additionally simplify some of our proofs. However, we use internal information, as this gives us the strongest information cost trade-off result.

Consider a randomized two-party communication protocol Π which uses public randomness R, and may additionally use private randomness. Suppose that M is the message transcript of the protocol, when the inputs to the two players, Alice and Bob, respectively, are sampled from the joint distribution λ. Let the input random variables be denoted by X, Y. The *information cost* of the protocol for Alice with respect to the distribution λ is defined as $\mathrm{IC}^{\mathsf{A}}_{\lambda}(\Pi) \stackrel{\mathrm{def}}{=} \mathrm{I}(X : M \,|\, YR)$. The information cost of the protocol for Bob is defined symmetrically as $\mathrm{IC}^{\mathsf{B}}_{\lambda}(\Pi) \stackrel{\mathrm{def}}{=} \mathrm{I}(Y : M \,|\, XR)$. These quantities measure the amount of information about one party's input that the other gains through the course of the protocol.

Note that we could have conditioned on the private randomness used by one party (say, Bob) as well in the other's (Alice's) information cost. This is however redundant, as given

his input Y, the public randomness R, and the message transcript M, Bob's private randomness R_{B} is independent of Alice's input (and private randomness). Indeed, by the Chain Rule (Proposition 2.3),

$$\begin{aligned} \mathrm{I}(X : M \,|\, YRR_{\mathsf{B}}) &= \mathrm{I}(X : MR_{\mathsf{B}} \,|\, YR) - \mathrm{I}(X : R_{\mathsf{B}} \,|\, YR) \\ &= \mathrm{I}(X : MR_{\mathsf{B}} \,|\, YR) \\ &= \mathrm{I}(X : M \,|\, YR) + \mathrm{I}(X : R_{\mathsf{B}} \,|\, YRM) \\ &= \mathrm{I}(X : M \,|\, YR) \ . \end{aligned}$$

2.3 The classical information cost lower bound

The first main theorem in this article may be viewed as a trade-off between information revealed by the two parties about their inputs while computing the Augmented Index function f_n. We show that at least one of the parties necessarily reveals "a lot" of information even on an "easy distribution" if the protocol computes f_n with bounded error on a "hard distribution".

Recall that in the Augmented Index problem, one party, Alice, has an n-bit string x, and the other party, Bob, has an integer $k \in [n]$, the prefix $x[1, k-1]$ of x, and a bit $b \in \{0, 1\}$. Their goal is to compute the function $f_n(x, (k, x[1, k-1], b)) = x_k \oplus b$, i.e., to determine whether $b = x_k$ or not, by engaging in a two-party communication protocol.

Let (X, K, B) be random variables distributed according to μ, the uniform distribution over $\{0, 1\}^n \times [n] \times \{0, 1\}$. Let μ_0 denote the distribution conditioned upon $B = X_K$, i.e., when the inputs are chosen uniformly from the set of 0s of f_n. We are interested in the information cost of a protocol Π with public randomness R for Augmented Index under the distribution μ_0, for the two parties. Let M denote the entire message transcript under μ, and let M^0 denote the transcript under distribution μ_0. Then the information cost of Π is given by $\mathrm{IC}^{\mathsf{A}}_{\mu_0}(\Pi) = \mathrm{I}(X : M^0 \,|\, X[1, K]\, R)$ and $\mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi) = \mathrm{I}(K : M^0 \,|\, XR)$. Note that $X[1, K] = X[1, K-1]B$ under distribution μ_0 and that K can be computed from $X[1, K]$. Hence K, B are not explicitly included in Bob's input in the expression for Alice's information cost. Similarly, $X[1, K-1]\, B$ are determined by K when we condition on X under distribution μ_0. Hence, these are not explicitly included in Bob's input in the expression for his information cost. The use of the notation M^0 is equivalent to conditioning on the event $X_K = B$, i.e., imposing the distribution μ_0, and helps us present our arguments more cleanly.

Since the value of the Augmented Index function f_n is a constant on μ_0, there is no *a priori* reason for the information cost of any party in a protocol to be large. However, we additionally require the protocol to be correct with non-trivial probability on the uniform distribution, under which there is equal chance of the function being 0 or 1. If the information cost (under μ_0) of the two parties is sufficiently low, we show that neither party can determine with high enough confidence what the function value is. The intuition behind this is as follows. Suppose we restrict the inputs to μ_0. If Bob's input K is changed, the random variables in Alice's possession, specifically the message transcript M^0 conditioned on her inputs, are not perturbed by much. This is because these random variables reveal little information about K. Similarly, if we flip one of the bits of Alice's input X outside of the prefix with Bob, the random variables in Bob's possession at the end of the protocol are not perturbed by much. Formally, these properties follow from the Average Encoding Theorem. Observe that if we simultaneously change Bob's index K to some $L > K$ and flip the Lth bit of X, we switch from a 0-input of f_n to a 1-input. The Cut-and-Paste Lemma ensures that by simultaneously changing the inputs with the two parties, the message transcript is perturbed by at most the sum of the amounts when the inputs are changed one at a time. This implies that the message transcript does not sufficiently help either party compute the function value.

We formalize this intuition in the next theorem, which we state for even n. A similar result holds for odd n, and may be derived from the proof for the even case. Together, they give us Theorem 1.2, as stated in the introduction (Section 1).

Theorem 2.6. *For any two-party randomized communication protocol* Π *for the* AUGMENTED INDEX *function* f_n *with* n *even, that makes error at most* $\varepsilon \in [0, 1/4)$ *on the uniform distribution* μ *over inputs, we have*

$$\left[\frac{\mathrm{IC}^{\mathsf{A}}_{\mu_0}(\Pi)}{n}\right]^{1/2} + \left[2 \cdot \mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi)\right]^{1/2} \quad \geq \quad \frac{1-4\varepsilon}{4\sqrt{\ln 2}} - \left[\frac{\mathrm{H}(2\varepsilon)}{n}\right]^{1/2} \ ,$$

where μ_0 *is the uniform distribution over* $f_n^{-1}(0)$*. In particular, for any* ε *smaller than* $1/4$ *by a constant, either* $\mathrm{IC}^{\mathsf{A}}_{\mu_0}(\Pi) \in \Omega(n)$ *or* $\mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi) \in \Omega(1)$*.*

Proof: Consider a protocol Π as in the statement of the theorem. Let the inputs be given by random variables X, K, B, drawn from the distribution μ.

Let M be the entire message transcript of the protocol, and let M^0 be the transcript under distribution μ_0. Without loss of generality, we assume that Bob computes the output of the protocol. If Alice computes the output, we include an additional message from her to Bob consisting of the output. We show below that this only marginally increases the information revealed by Alice, and include its effect in the lower bound we derive. Indeed, if the single bit output of the protocol is O^0 under the distribution μ_0, $\mathrm{H}(O^0) \leq \mathrm{H}(2\varepsilon)$, as the protocol produces the correct output with probability at least $1 - 2\varepsilon$ on the distribution μ_0. Let $d \geq 0$ be such that $\mathrm{I}(X : M^0 \,|\, X[1, K]) = dn$. Then,

$$\begin{aligned} \mathrm{I}(X : M^0 O^0 \,|\, X[1, K]) &= \mathrm{I}(X : M^0 \,|\, X[1, K]) + \mathrm{I}(X : O^0 \,|\, M^0 X[1, K]) \\ &\leq dn + \mathrm{H}(O^0) \ , \end{aligned}$$

and $\mathrm{I}(K : M^0 O^0 \,|\, X) = \mathrm{I}(K : M^0 \,|\, X)$. Henceforth, we assume that the output of the protocol Π is computed by Bob, and its information costs are bounded as $\mathrm{IC}^{\mathsf{A}}_{\mu_0}(\Pi) \leq d_1 n$ with $d_1 = d + \mathrm{H}(2\varepsilon)/n$, and $\mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi) \leq c$.

Let R be the public randomness used in the protocol. For each specific value r for the public random coins, we use the subscript r on a random variable to denote conditioning on $R = r$. In particular, the random variable M^0_r is the transcript M conditioned on $R = r$, under distribution μ_0. Define $d_{1r} \stackrel{\text{def}}{=} \frac{1}{n}\ \mathrm{I}(X : M^0_r \,|\, X[1, K])$ and $c_r \stackrel{\text{def}}{=} \mathrm{I}(K : M^0_r \,|\, X)$, so that $\mathbb{E}_{r \leftarrow R}\ d_{1r} = \mathrm{IC}^{\mathsf{A}}_{\mu_0}(\Pi)/n$ and $\mathbb{E}_{r \leftarrow R}\ c_r = \mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi)$. We emphasize that the protocol may use private randomness in addition to the public randomness R. Let ε_r denote the error made by the protocol Π on the uniform distribution μ over inputs, when $R = r$.

In the rest of the proof, we fix a specific value r for the public randomness, and show that

$$d_{1r}^{1/2} + (2c_r)^{1/2} \quad \geq \quad \frac{1-4\varepsilon_r}{4\sqrt{\ln 2}} \ . \tag{2.1}$$

Averaging this over $r \leftarrow R$ and applying the Jensen Inequality gives us the theorem.

We show below that the random variables $M^0_r X[1, K]$ with Bob are "close" in distribution to the random variables $M^1_r X[1, K-1]\, \bar{X}_K$, where M^1_r denotes the transcript M_r conditioned on the function value being 1, i.e., when $B = \bar{X}_K$. In other words, we show that the ℓ_1 distance between them is only "slightly more" than 1 if the information cost of the protocol is small.

Lemma 2.7. $\left\| M^0_r X[1, K] - M^1_r X[1, K-1]\, \bar{X}_K \right\| \quad \leq \quad 1 + 8\sqrt{\kappa\, c_r} + 4\sqrt{2\kappa\, d_{1r}}$*, where* $\kappa = \frac{\ln 2}{2}$*.*

For any fixed r, given the message transcript and his input, Bob's private randomness is independent of Alice's input and private randomness. Therefore, we can regenerate Bob's private randomness exactly from the other random variables in his possession. As a result, we may use the protocol Π to identify the two distributions, $M^0_r X[1, K]$ and $M^1_r X[1, K-1]\, \bar{X}_K$, with average error ε_r. If the error ε_r were small, the ℓ_1 distance would be correspondingly closer to 2. Formally, the ℓ_1 distance between two distributions is non-increasing under the action of a stochastic map. So $\left\| M^0_r X[1, K] - M^1_r X[1, K-1]\, \bar{X}_K \right\| \geq 2(1 - 2\varepsilon_r)$, as the latter is a lower bound on the ℓ_1 distance between the distributions of the output of the protocol in the two cases. This gives us a lower bound on the information cost, in terms of the error made by the

protocol. Combining the two bounds on the ℓ_1 distance, we get Eq. (2.1) and hence the theorem. ■

We now prove the heart of the theorem, i.e., that the message transcript for the 0 and 1 inputs are close to each other in distribution.

Proof of Lemma 2.7: The proof follows the intuition given before Theorem 2.6. We break the proof into several steps, each of which is captured by a lemma. The proofs of the lemmata are postponed to later in the section so as to present the high-level argument first.

When we wish to explicitly write the transcript M_r as a function of the inputs to Alice and Bob, say x and $x[1,k-1], b$ respectively, we write it as $M_r(x; x[1,k-1], b)$. If $b = x_k$, we write Bob's input as $x[1,k]$.

For any $x \in \{0,1\}^n$ and $i \in [n]$, let $x^{(i)}$ denote the string that equals x in all coordinates except at the ith. Since $(X, X[1,K-1], \bar{X}_K)$ and $(X^{(K)}, X[1,K])$ are identically distributed, $M_r^1 = M_r(X; X[1,K-1], \bar{X}_K)$ has the same distribution as $M_r(X^{(K)}; X[1,K])$. Thus, our goal is to bound

$$\Big\| M_r(X; X[1,K])\, X[1,K] - M_r(X^{(K)}; X[1,K])\, X[1,K] \Big\| \quad .$$

Later, we consider the random variables in Bob's possession when we flip one of the bits in input X with Alice. In order to do the flip in a manner consistent with the prefix with Bob, we only flip bits in coordinates $> n/2$. This gives us a bound on the above quantity when the index is larger than $n/2$. Therefore we consider L uniformly and independently distributed in $[n] - [n/2]$, and J be uniformly and independently distributed in $[n/2]$. We have

$$\begin{aligned}
&\Big\| M_r(X; X[1,K])\, X[1,K] - M_r(X^{(K)}; X[1,K])\, X[1,K] \Big\| \\
&= \Big\| \frac{1}{2} \left(M_r(X; X[1,J])\, X[1,J] + M_r(X; X[1,L])\, X[1,L] \right) \\
&\quad - \frac{1}{2} \Big(M_r(X^{(J)}; X[1,J])\, X[1,J] + M_r(X^{(L)}; X[1,L])\, X[1,L] \Big) \Big\| \\
&\leq \frac{1}{2} \Big\| M_r(X; X[1,J])\, X[1,J] - M_r(X^{(J)}; X[1,J])\, X[1,J] \Big\| \\
&\quad + \frac{1}{2} \Big\| M_r(X; X[1,L])\, X[1,L] - M_r(X^{(L)}; X[1,L])\, X[1,L] \Big\| \\
&\leq 1 + \frac{1}{2} \Big\| M_r(X; X[1,L])\, X[1,L] - M_r(X^{(L)}; X[1,L])\, X[1,L] \Big\| \quad , \qquad (2.2)
\end{aligned}$$

and we bound the RHS from above.

Recall that our goal is to show that, on average, changing from a 0-input to a 1-input does not perturb the message transcript by much. For this, we begin by showing that changing Alice's input alone, or similarly, Bob's input alone, has this kind of effect. If the information cost of Bob is small, the message transcript does not carry much information about K when the inputs are drawn from μ_0. From this, we deduce that the transcript M_r^0 is (on average) nearly the same for different inputs to Bob.

We compare the transcript when Bob's input index is J to when it is L.

Lemma 2.8. $\mathbb{E}_{(x,j,l) \leftarrow (X,J,L)}\, \mathfrak{h}(M_r(x\,;\, x[1,j])\ ,\ M_r(x\,;\, x[1,l]))^2 \quad \leq \quad 8\kappa\, c_r.$

We defer the proof to later in this section.

In the interest of readability, we abbreviate some random variables in the rest of the proof, as also in the intermediate lemmata. For $i \in [n]$, and a prefix $x[1,i]$ of a string $x \in \{0,1\}^n$ that will be clear from the context, let v_i denote the prefix $x[1,i]$, let U_i denote the random variable $x[1,i]\, X[i+1,n]$ (i.e., X conditioned on having prefix v_i), and let U_i' denote the random variable $x[1,i-1]\, \bar{x}_i\, X[i+1,n]$ (i.e., U_i with the ith bit flipped).

When changing Alice's input, we would like to ensure that the prefix held by Bob does not change. So we restrict our attention to Bob's inputs with index $J \in [n/2]$, and change Alice's input by flipping the Lth bit, with $L \in [n] - [n/2]$. If the information cost of Alice is small, M_r^0

does not carry much information about X, even given a prefix. Therefore, flipping a bit outside the prefix does not perturb the transcript by much.

Lemma 2.9. $\mathbb{E}_{(x[1,l],j,l)\leftarrow(X[1,L],J,L)}\ \mathfrak{h}(M_r(U_l\,;\,v_j)\ ,\ M_r(U'_l\,;\,v_j))^2 \quad\leq\quad 16\kappa\, d_{1r}\ .$

This is proven later in the section.

We now conclude the proof of Lemma 2.7. Since Hellinger distance squared is jointly convex (Proposition 2.2), Lemma 2.8 gives us a bound on the distance between the transcripts averaged over the choice of suffix $x[l+1,n]$. Along with the Jensen Inequality, we get

$$\mathbb{E}_{(x[1,l],j,l)\leftarrow(X[1,L],J,L)}\ \mathfrak{h}(M_r(U_l\,;\,v_j)\ ,\ M_r(U_l\,;\,v_l)) \quad\leq\quad \sqrt{8\kappa\, c_r}\ . \tag{2.3}$$

Along with the Triangle Inequality, and Lemma 2.9, this implies that

$$\mathbb{E}_{(x[1,l],j,l)\leftarrow(X[1,L],J,L)}\ \mathfrak{h}(M_r(U_l\,;\,v_l)\ ,\ M_r(U'_l\,;\,v_j)) \quad\leq\quad \sqrt{8\kappa\, c_r}+\sqrt{16\kappa\, d_{1r}}\ .$$

Using the Cut-and-Paste property of private coin communication protocols (Proposition 2.5), we conclude that simultaneously changing Bob's input from $x[1,j]$ to $x[1,l]$ and flipping the lth bit of x perturbs the transcript by no more than the individual changes.

$$\begin{aligned}\mathbb{E}_{(x[1,l],j,l)\leftarrow(X[1,L],J,L)}\ &\mathfrak{h}(M_r(U_l\,;\,v_j)\ ,\ M_r(U'_l\,;\,v_l))\\ &=\quad \mathbb{E}_{(x[1,l],j,l)\leftarrow(X[1,L],J,L)}\ \mathfrak{h}(M_r(U_l\,;\,v_l)\ ,\ M_r(U'_l\,;\,v_j))\\ &\leq\quad \sqrt{8\kappa\, c_r}+\sqrt{16\kappa\, d_{1r}}\ .\end{aligned} \tag{2.4}$$

Combining Eq. (2.3) and Eq. (2.4), and using the Triangle Inequality we get

$$\mathbb{E}_{(x[1,l],l)\leftarrow(X[1,L],L)}\ \mathfrak{h}(M_r(U_l\,;\,v_l)\ ,\ M_r(U'_l\,;\,v_l)) \quad\leq\quad 4\sqrt{2\kappa\, c_r}+4\sqrt{\kappa\, d_{1r}}\ .$$

Using Proposition 2.1, we translate this back to a bound on ℓ_1 distance:

$$\begin{aligned}\Big\|M_r(X\,;\,X[1,L])\,X[1,L]\ -\ M_r(X^{(L)}\,;\,X[1,L])\,X[1,L]\Big\|&\\ \leq\quad \mathbb{E}_{(x[1,l],l)\leftarrow(X[1,L],L)}\,\|M_r(U_l\,;\,v_l)-M_r(U'_l\,;\,v_l)\|&\\ \leq\quad 16\sqrt{\kappa\, c_r}+8\sqrt{2\kappa\, d_{1r}}\ .&\end{aligned}$$

Lemma 2.7 follows by combining this with Eq. (2.2). ∎

We return to the lemmata whose proofs we had deferred.

Lemma 2.8. $\mathbb{E}_{(x,j,l)\leftarrow(X,J,L)}\ \mathfrak{h}(M_r(x\,;\,x[1,j])\ ,\ M_r(x\,;\,x[1,l]))^2 \quad\leq\quad 8\kappa\, c_r.$

Proof: Let us define a new random variable $\tilde{M}_r$ jointly distributed with X, and independent of all other random variables, such that the joint distribution of $X\tilde{M}_r$ is identical to the joint distribution of XM_r^0. In particular, we have $\tilde{M}_r(x)=\mathbb{E}_{k\leftarrow K}M_r(x\,;\,x[1,k])$.

By the Average Encoding Theorem, Proposition 2.4, we have that for every $x\in\{0,1\}^n$,

$$\mathbb{E}_{k\leftarrow K}\ \mathfrak{h}\Big(M_r(x\,;\,x[1,k])\ ,\ \tilde{M}_r(x)\Big)^2 \quad\leq\quad \kappa\,\mathrm{I}(K:M_r^0\,|\,X=x)\ ,$$

where $\kappa=\frac{\ln 2}{2}$. Averaging over $x\leftarrow X$,

$$\mathbb{E}_{(x,k)\leftarrow(X,K)}\ \mathfrak{h}\Big(M_r(x\,;\,x[1,k])\ ,\ \tilde{M}_r(x)\Big)^2 \quad\leq\quad \kappa\,\mathrm{I}(K:M_r^0\,|\,X) \quad=\quad \kappa\, c_r\ .$$

An immediate consequence is that

$$\begin{aligned}\mathbb{E}_{(x,j)\leftarrow(X,J)}\ \mathfrak{h}\Big(M_r(x\,;\,x[1,j])\ ,\ \tilde{M}_r(x)\Big)^2 \quad&\leq\quad 2\,\kappa\, c_r\ , \qquad\text{and}\\ \mathbb{E}_{(x,l)\leftarrow(X,L)}\ \mathfrak{h}\Big(M_r(x\,;\,x[1,l])\ ,\ \tilde{M}_r(x)\Big)^2 \quad&\leq\quad 2\,\kappa\, c_r\ .\end{aligned}$$

By the Triangle Inequality, for any $j \in [n/2]$, $l \in [n] - [n/2]$, and $x \in \{0,1\}^n$,

$$\begin{aligned}
&\mathfrak{h}(M_r(x\,;\,x[1,j])\;,\;M_r(x\,;\,x[1,l]))^2 \\
&\quad\leq\quad \Big(\mathfrak{h}\Big(M_r(x\,;\,x[1,j])\;,\;\tilde{M}_r(x)\Big) + \mathfrak{h}\Big(M_r(x\,;\,x[1,l])\;,\;\tilde{M}_r(x)\Big)\Big)^2 \\
&\quad\leq\quad 2\,\mathfrak{h}\Big(M_r(x\,;\,x[1,j])\;,\;\tilde{M}_r(x)\Big)^2 + 2\,\mathfrak{h}\Big(M_r(x\,;\,x[1,l])\;,\;\tilde{M}_r(x)\Big)^2 \quad .
\end{aligned}$$

Taking expectation over X, J, L, we get the claimed bound. ■

Lemma 2.9. $\mathbb{E}_{(x[1,l],j,l)\leftarrow(X[1,L],J,L)}\;\mathfrak{h}(M_r(U_l\,;\,v_j)\;,\;M_r(U'_l\,;\,v_j))^2 \quad\leq\quad 16\kappa\,d_{1r}$.

Proof: This intuition behind this lemma is the same as that behind the impossibility of "random access encoding" [40, 3], as we explain next. Suppose we view the transcript as an encoding of the bits of X not known to Bob, of which there are at least $n/2$. Since they are uniformly random, the net information in the encoding about the bits is no more than the sum of the information about the individual bits, even conditioned on the prefix. This follows by the superadditivity of mutual information for independent random variables (equivalently, the Chain Rule, Proposition 2.3). This implies that, on average, the encoding is very weakly correlated with the bits. The Average Encoding Theorem (Proposition 2.4) then implies that the messages for two prefixes that differ in one bit are close to each other, on average. We formalize this below.

We have

$$\begin{aligned}
d_{1r}n \quad&\geq\quad \mathrm{I}(X : M_r^0 \,|\, X[1,K]) \\
&=\quad \frac{1}{2}\,\mathbb{E}_{j\leftarrow J}\,\mathrm{I}(X : M_r(X\,;\,X[1,J])\,|\,X[1,J]) + \frac{1}{2}\,\mathbb{E}_{l\leftarrow L}\,\mathrm{I}(X : M_r(X\,;\,X[1,L])\,|\,X[1,L]) \\
&\geq\quad \frac{1}{2}\,\mathbb{E}_{j\leftarrow J}\,\mathrm{I}(X : M_r(X\,;\,X[1,J])\,|\,X[1,J]) \quad . \qquad (2.5)
\end{aligned}$$

Fix a sample point $(x[1,j], j)$, with $j \in [n/2]$. By the Chain Rule (Proposition 2.3),

$$\begin{aligned}
&\mathrm{I}(X[j+1,n] : M_r(U_j\,;\,v_j)) \qquad (2.6)\\
&\quad=\quad \sum_{l=j+1}^{n} \mathrm{I}(X_l : M_r(U_j\,;\,v_j)\,|\,X[j+1,l-1]) \\
&\quad\geq\quad \sum_{l=n/2+1}^{n} \mathrm{I}(X_l : M_r(U_j\,;\,v_j)\,|\,X[j+1,l-1]) \quad . \qquad (2.7)
\end{aligned}$$

Moreover, by the Triangle Inequality and the Average Encoding Theorem (Proposition 2.4), for any given $x[1,l]$, with $l \in [n] - [n/2]$,

$$\begin{aligned}
&\mathfrak{h}(M_r(U_l\,;\,v_j)\;,\;M_r(U'_l\,;\,v_j))^2 \\
&\leq\quad \big[\;\mathfrak{h}(M_r(U_l\,;\,v_j)\;,\;M_r(U_{l-1}\,;\,v_j)) + \mathfrak{h}(M_r(U'_l\,;\,v_j)\;,\;M_r(U_{l-1}\,;\,v_j))\;\big]^2 \\
&\leq\quad 2\,\big[\;\mathfrak{h}(M_r(U_l\,;\,v_j)\;,\;M_r(U_{l-1}\,;\,v_j))^2 + \mathfrak{h}(M_r(U'_l\,;\,v_j)\;,\;M_r(U_{l-1}\,;\,v_j))^2\;\big] \\
&\leq\quad 4\kappa\,\mathrm{I}(X_l : M_r(U_{l-1}\,;\,v_j)) \quad . \qquad (2.8)
\end{aligned}$$

Combining Eqs. (2.5), (2.7), and (2.8), we get

$$\begin{aligned}
&\mathbb{E}_{(x[1,l],j,l)\leftarrow(X[1,L],J,L)}\;\mathfrak{h}(M_r(U_l\,;\,v_j)\;,\;M_r(U'_l\,;\,v_j))^2 \\
&\leq\quad 4\kappa\,\mathbb{E}_{(x[1,l-1],j,l)\leftarrow(X[1,L-1],J,L)}\;\mathrm{I}(X_l : M_r(U_{l-1}\,;\,v_j)) \\
&=\quad 4\kappa\,\mathbb{E}_{(x[1,j],j,l)\leftarrow(X[1,J],J,L)}\;\mathrm{I}(X_l : M_r(U_j\,;\,v_j)\,|\,X[j+1,l-1]) \\
&\leq\quad \frac{8\kappa}{n}\,\mathrm{I}(X : M_r(X\,;\,X[1,J])\,|\,X[1,J]) \quad\leq\quad 16\kappa\,d_{1r} \quad ,
\end{aligned}$$

as claimed. ■

3 The connection with streaming algorithms

Streaming algorithms are algorithms of a simple form, intended to process massive problem instances rapidly, ideally using space that is of smaller order than the size of the input. A *pass* on an input $x \in \Sigma^n$, where Σ is some alphabet, means that x is read as an *input stream* $x_1, x_2, \ldots, x_n$, which arrives sequentially, i.e., letter by letter in this order.

Definition 3.1 (Streaming algorithm)**.** *Fix an alphabet Σ. A (unidirectional) T-pass* streaming algorithm $\boldsymbol{A}$ *with space $s(n)$ and time $t(n)$ is an algorithm such that for every input stream $x \in \Sigma^n$:*

1. $\boldsymbol{A}$ *performs T sequential passes on x in the order $x_1, x_2, \ldots, x_n$,*
2. $\boldsymbol{A}$ *maintains a memory space of size $s(n)$ bits while reading x,*
3. $\boldsymbol{A}$ *has running time at most $t(n)$ per letter x_i, and*
4. $\boldsymbol{A}$ *has pre-processing and post-processing time at most $t(n)$.*

We say that $\boldsymbol{A}$ is bidirectional *if it is allowed to read the input in the reverse order, after reaching the last letter. Then the parameter T is the total number of passes in either direction.*

In general, the pre- and post-processing times of a streaming algorithm may be different, and may differ from the running time per letter. Since the results in this section apply to streaming algorithms regardless of their time complexity, we choose not to make this finer distinction.

We refer the reader to the text [39] for a more thorough introduction to streaming algorithms.

Recall that in a two-party communication protocol for Augmented Index, one party, Alice, has an n-bit string x, and the other party, Bob, has an integer $k \in [n]$, the prefix $x[1, k-1]$ of x, and a bit $b \in \{0, 1\}$. Their goal is to compute the function $f_n(x, (k, x[1, k-1], b)) = x_k \oplus b$, i.e., to determine whether $b = x_k$ or not, by engaging in a two-party communication protocol.

The relationship between streaming algorithms for Dyck(2) and communication protocols for f_n is captured by a reduction due to Magniez, Mathieu, and Nayak [37]. The reduction was originally described only for one-pass streaming algorithms, but extends readily to unidirectional multi-pass algorithms. For completeness, we include a proof of this theorem here.

Theorem 3.1. *Suppose there is a randomized unidirectional streaming algorithm for* Dyck(2) *with T passes that uses space s for instances of length at most $4n^2$, and has worst-case two-sided error δ. Then there is a two-party communication protocol Π for the* Augmented Index *function f_n that makes error at most δ on the uniform distribution μ over its inputs, and has information costs $\mathrm{IC}^{\mathsf{A}}_{\mu_0}(\Pi) \leq sT$ for Alice and $\mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi) \leq sT/n$ for Bob, with respect to the uniform distribution μ_0 over $f_n^{-1}(0)$.*

Proof: For any string $z = z_1 \cdots z_n \in \{a, b\}^n$, let $\overline{z}$ denote the matching string $\overline{z_n}\,\overline{z_{n-1}} \cdots \overline{z_1}$ corresponding to z. Let $z[i, j]$ denote the substring $z_i z_{i+1} \cdots z_j$ if $1 \leq i \leq j \leq n$, and the empty string ϵ otherwise. We abbreviate $z[i, i]$ as $z[i]$ if $1 \leq i \leq n$.

We focus on a subset of instances for Dyck(2) defined as follows. Let n be a positive integer. Consider strings of the form

$$w \;=\; x^1\,\overline{y^1}\,\overline{z^1}\,z^1\,y^1\;x^2\,\overline{y^2}\,\overline{z^2}\,z^2\,y^2\;\cdots\;x^n\,\overline{y^n}\,\overline{z^n}\,z^n\,y^n\;\overline{x^n}\;\cdots\;\overline{x^2}\;\overline{x^1}\;, \tag{3.1}$$

where for every i, $x^i \in \{0, 1\}^n$, $y^i = x^i[n - k^i + 2, n]$ for some $k^i \in \{1, 2, \ldots, n\}$, and $z^i \in \{a, b\}$. The string w is in Dyck(2) if and only if, for every i, $z^i = x^i[n - k^i + 1]$. Note that these instances have length in the interval $[2n(n+1), 4n^2]$. Figure 1 depicts an instance of this form.

Intuitively, recognizing strings of the form w is difficult in one pass with space $o(n)$. After reading x^i, the streaming algorithm does not have enough space to store this string so as to be able to check the bit at unknown index $(n - k_i + 1)$. Moreover, after reading $\overline{y_n}$ it does not have enough space to store information about all indices $k^1, k^2, \ldots, k^n$. When it reads $\overline{x^n} \cdots \overline{x^2}\,\overline{x^1}$ it therefore misses out on its second chance to check whether $z^i = x^i[n - k^i + 1]$ for every i. When the algorithm is allowed a larger number of passes T in the same direction, it may adopt a

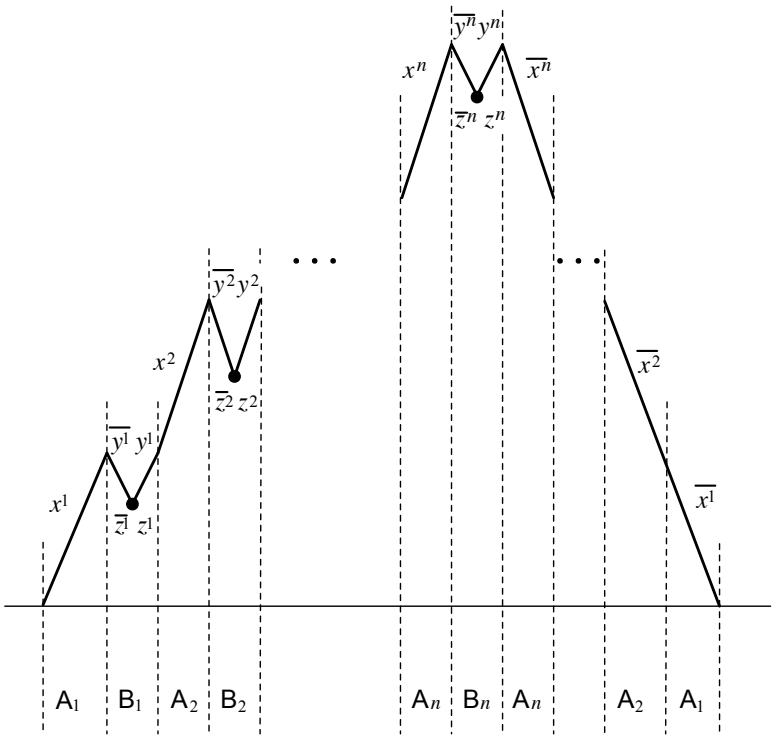

Figure 1: An instance of the form described in Eq. (3.1). A line segment with positive slope denotes a string over $\{a, b\}$, and a segment with negative slope denotes a string over $\{\overline{a}, \overline{b}\}$. A solid dot depicts a pair of the form $\overline{z}z$ for some $z \in \{a, b\}$. The entire string is distributed amongst $2n$ players $\mathsf{A}_1, \mathsf{B}_1, \mathsf{A}_2, \mathsf{B}_2, \ldots, \mathsf{A}_n, \mathsf{B}_n$ in a communication protocol for Ascension(n) as shown.

more sophisticated strategy. Nevertheless, the same intuition carries over with a tighter bound of $o(n/T)$ on the space.

We observe that a space s streaming algorithm gives rise to a multiparty communication protocol for the problem Ascension(n), which is the logical OR of n independent instances of the Augmented Index function f_n. In more detail, in the problem Ascension(n) there are $2n$ players $\mathsf{A}_1, \mathsf{A}_2, \ldots, \mathsf{A}_n$ and $\mathsf{B}_1, \mathsf{B}_2, \ldots, \mathsf{B}_n$. Player A_i is given $x^i \in \{0,1\}^n$, player B_i is given $k^i \in [n]$, a bit z^i, and the prefix $x^i[1, k^i - 1]$ of x^i. Let $\mathbf{x} = (x^1, x^2, \ldots, x^n)$, $\mathbf{k} = (k^1, k^2, \ldots, k^n)$, and $\mathbf{z} = (z^1, z^2, \ldots, z^n)$.

The goal of the communication protocol is to compute

$$F_n(\mathbf{x}, \mathbf{k}, \mathbf{z}) \quad = \quad \bigvee_{i=1}^{n} f_n(x^i, k^i, z^i) \quad = \quad \bigvee_{i=1}^{n} (x^i[k^i] \oplus z^i) \ ,$$

which is 0 if $x^i[k^i] = z^i$ for all i, and 1 otherwise. The communication between the $2n$ parties is required to be T sequential iterations of communication in the following order, for some $T \geq 1$:

$$\mathsf{A}_1 \to \mathsf{B}_1 \to \mathsf{A}_2 \to \mathsf{B}_2 \to \cdots \mathsf{A}_n \to \mathsf{B}_n \to \mathsf{A}_n \to \mathsf{A}_{n-1} \to \cdots \to \mathsf{A}_2 \to \mathsf{A}_1 \ . \tag{3.2}$$

In other words, for $t = 1, 2, \ldots, T$,

- for i from 1 to $n-1$, player A_i sends message $M_{\mathsf{A}_i,t}$ to B_i, then B_i sends message $M_{\mathsf{B}_i,t}$ to A_{i+1},
- A_n sends message $M_{\mathsf{A}_n,t}$ to B_n,
- B_n sends message $M_{B_n,t}$ to A_n,

– for i from n down to 2, A_i sends message $M'_{\mathsf{A}_i,t}$ to A_{i-1}.

At the end of the T iterations, A_1 computes the output.

There is a one-to-one correspondence between inputs to DYCK(2) of the form in Eq. (3.1) and the inputs to ASCENSION(n). This arises from a partition of the word among $2n$ players as depicted in Figure 1. For ease of notation, the strings x^i in ASCENSION(n) are taken to be the ones in DYCK(2) with the bits *in reverse order*. This switches the suffixes y^i with prefixes of the same length.

The following is immediate.

Lemma 3.2. *A unidirectional T-pass streaming algorithm for* DYCK(2) *with space s implies a communication protocol for* ASCENSION(n) *with T iterations of communication as above, in which every message is of length s. Moreover, on any input, the probability of error of the protocol is the same as that of the algorithm.*

Proof: In each of the T iterations, a player simulates the streaming algorithm on his/her part of the input, and sends the length s workspace to the next player in the sequence. The final player A_1 gives the output of the algorithm as that of the protocol. ■

We prove a *direct sum* result that captures the relationship of ASCENSION(n) to solving n instances of the more "primitive" problem AUGMENTED INDEX. The direct sum result is proven using the superadditivity of mutual information for inputs (x^i, k^i, z^i) picked independently from the uniform distribution μ_0 over $f_n^{-1}(0)$. The use of this "easy" distribution collapses the function ASCENSION(n) to an instance of AUGMENTED INDEX in any chosen coordinate. The direct sum result allows us to choose a coordinate with small information cost, which proves the theorem.

Consider an instance $(\mathbf{X}, \mathbf{K}, \mathbf{Z})$ of ASCENSION(n) distributed according to μ_0^n over $(\{0,1\}^n \times [n] \times \{0,1\})^n$, where $\mathbf{X} = (X^1, X^2, \ldots, X^n)$, $\mathbf{K} = (K^1, K^2, \ldots, K^n)$ and $\mathbf{Z} = (Z^1, Z^2, \ldots, Z^n)$.

Let $\tilde{\Pi}$ be a public-coin randomized protocol for ASCENSION(n) derived from a unidirectional T-pass streaming algorithm for DYCK(2). Assume it has worst-case error δ, and that each message is of length at most s. For each $j \in [n]$, we construct a protocol Π_j as follows for the AUGMENTED INDEX function f_n. Let (x, k, c) be the input for AUGMENTED INDEX.

1. Alice sets A_j's input x^j to her input x.
2. Bob sets B_j's input $(k^j, x^j[1, k^j - 1], z^j)$ to his input $(k, x[1, k-1], c)$.
3. Alice and Bob generate, using public coins, X^i uniformly at random from $\{0,1\}^n$, independently for all $i > j$, and (X^i, K^i, Z^i) distributed according to μ_0, independently for all $i < j$.
4. Bob generates K^i uniformly and independently for $i > j$, using private coins. Then Bob sets $Z^i = X^i[k^i]$ for $i > j$, so that (X^i, K^i, Z^i) are distributed according to μ_0, independently for all $i > j$.
5. Alice and Bob simulate the protocol $\tilde{\Pi}$ by executing the roles of players $(\mathsf{A}_i, \mathsf{B}_i)_{i=1}^n$ as follows. In the tth iteration of communication in the order described in Eq. (3.2),
   (a) Alice runs $\tilde{\Pi}$ until she generates the message $M_{\mathsf{A}_j,t}$ from player A_j. She sends this message to Bob.
   (b) Bob continues running $\tilde{\Pi}$ until he generates the message $M_{\mathsf{B}_n,t}$ from player B_n. He sends this message to Alice.
   (c) Alice completes the rest of the tth iteration of $\tilde{\Pi}$ until she generates the message $M'_{\mathsf{A}_2,t}$ from player A_2, and moves to the next iteration of $\tilde{\Pi}$ (if any).

   At the end of the Tth iteration, Alice completes the rest of the protocol $\tilde{\Pi}$ and produces as output for Π_j, the output of player A_1 in $\tilde{\Pi}$.

By definition of the distribution μ_0, we have $f_n(X^i, K^i, Z^i) = 0$ for all $i \neq j$. So $F_n(\mathbf{X}, \mathbf{K}, \mathbf{Z}) = f_n(x, k, c)$, and each protocol Π_j computes the function f_n, i.e., solves AUGMENTED INDEX, with worst-case error at most δ.

Note that in the simulation of $\tilde{\Pi}$ by Alice and Bob above, the random variables (X^i, K^i, Z^i) for $i < j$ are used only by Alice, and could have been generated by Alice using private coins. Making these random variables public does not affect the correctness of Π_j, but turns out to be convenient in deriving the direct sum result.

Let R denote the public coins used in the protocol $\tilde{\Pi}$. Let $\mathbf{M}$ denote the sequence of T random variables $M_{\mathsf{B}_n,1} M_{\mathsf{B}_n,2} \cdots M_{\mathsf{B}_n,T}$, viz., the messages sent by B_n over all the iterations. By the Chain Rule (Proposition 2.3),

$$\mathrm{I}(\mathbf{KZ} : \mathbf{M} \mid \mathbf{X}R) \quad = \quad \sum_{j=1}^{n} \mathrm{I}(K^j Z^j : \mathbf{M} \mid \mathbf{X} R\, K^1 Z^1 \cdots K^{j-1} Z^{j-1}) \ .$$

Let $R_j = (R, (X^i)_{j \neq i}, (K^i, Z^i)_{i<j})$. These are all the public random coins used in the protocol Π_j, and any further random coins are used only by Bob privately to generate $(K^i, Z^i)_{i>j}$. In particular, Alice does not use any private coins and her messages are (deterministic) functions of $X^j R_j$ and the messages received from Bob. Thus, for all j

$$\begin{aligned} \mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi_j) &= \mathrm{I}(K^j Z^j : \mathbf{M} \mid X^j R_j) \\ &= \mathrm{I}(K^j Z^j : \mathbf{M} \mid \mathbf{X} R\, K^1 Z^1 \cdots K^{j-1} Z^{j-1}) \ , \end{aligned}$$

and we have the direct sum result

$$\sum_{j=1}^{n} \mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi_j) \quad = \quad \mathrm{I}(\mathbf{KZ} : \mathbf{M} \mid \mathbf{X}R) \ .$$

Furthermore, $\mathbf{M}$ has length at most sT, so that

$$\sum_{j=1}^{n} \mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi_j) \quad \leq \quad sT \ ,$$

and there is a $j_0 \in [n]$ such that $\mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi_{j_0}) \leq sT/n$. We also have, by the Chain Rule (Proposition 2.3),

$$\begin{aligned} \mathrm{IC}^{\mathsf{A}}_{\mu_0}(\Pi_{j_0}) &= \mathrm{I}(X^{j_0} : M_{\mathsf{A}_{j_0},1} M_{\mathsf{A}_{j_0},2} \cdots M_{\mathsf{A}_{j_0},T}\, \mathbf{M} \mid K^{j_0} Z^{j_0}\, R_{j_0}) \\ &= \sum_{t=1}^{T} \big[\mathrm{I}(X^{j_0} : M_{\mathsf{A}_{j_0},t} \mid K^{j_0} Z^{j_0}\, R_{j_0} M_{\mathsf{A}_{j_0},1} M_{\mathsf{B}_n,1} \cdots M_{\mathsf{A}_{j_0},t-1} M_{\mathsf{B}_n,t-1}) \\ &\qquad + \mathrm{I}(X^{j_0} : M_{\mathsf{B}_n,t} \mid K^{j_0} Z^{j_0}\, R_{j_0} M_{\mathsf{A}_{j_0},1} M_{\mathsf{B}_n,1} \cdots M_{\mathsf{A}_{j_0},t-1} M_{\mathsf{B}_n,t-1} M_{\mathsf{A}_{j_0},t}) \big] \\ &= \sum_{t=1}^{T} \mathrm{I}(X^{j_0} : M_{\mathsf{A}_{j_0},t} \mid K^{j_0} Z^{j_0}\, R_{j_0} M_{\mathsf{A}_{j_0},1} M_{\mathsf{B}_n,1} \cdots M_{\mathsf{A}_{j_0},t-1} M_{\mathsf{B}_n,t-1}) \ , \qquad (3.3) \end{aligned}$$

since Bob's tth message is independent of Alice's input, conditioned on his input, the public randomness, and the transcript until Alice's tth message. Since the length of each message $M_{\mathsf{A}_{j_0},t}$ is bounded by s, Eq (3.3) implies

$$\mathrm{IC}^{\mathsf{A}}_{\mu_0}(\Pi_{j_0}) \quad \leq \quad sT \ .$$

The protocol Π_{j_0} is the protocol claimed by the theorem. ■

The information cost trade-off in Theorem 2.6 implies that any streaming algorithm that makes a "small" number of passes over the input requires a "large" amount of space.

Corollary 3.3. *Any randomized unidirectional T-pass streaming algorithm for* Dyck(2) *that has worst-case two-sided error $\delta < 1/4$ uses space at least*

$$\frac{\lfloor \sqrt{N}/2 \rfloor}{T} \times \frac{1}{3 + 2\sqrt{2}} \left[\frac{1 - 4\delta}{4\sqrt{\ln 2}} - \left(\frac{\mathrm{H}(2\delta)}{\lfloor \sqrt{N}/2 \rfloor} \right)^{1/2} \right]^2$$

on instances of length N.

4 Quantum information cost of Augmented Index

We now turn to quantum communication. We present the necessary background on quantum information theory in Section 4.1, and discuss quantum protocols and information cost in Section 4.2. In Section 4.3, we show how the notion of average encoding may be applied also to quantum protocols for Augmented Index. The analysis of quantum protocols for Augmented Index involves a number of additional additional subtleties, which are also described along the way.

4.1 Quantum information theory basics

We continue the use of capital letters to denote random variables. We see these as special cases of quantum states, which are trace one positive semi-definite matrices. Indeed, random variables may be viewed as quantum states that are diagonal in a canonical basis. Quantum states are also denoted by capital letters P, Q, etc.

The trace distance $\|A - B\|_{\mathrm{tr}}$ between two quantum states A, B over the same Hilbert space is the metric induced by the trace norm $\|M\|_{\mathrm{tr}} = \mathrm{Tr}\sqrt{M^\dagger M}$. The fidelity between the two states is defined as $\mathrm{F}(A, B) = \left\|\sqrt{A}\sqrt{B}\right\|_{\mathrm{tr}}$. The Bures distance $\mathfrak{h}(A\ ,\ B)$ between the states is a metric arising from fidelity, and is defined as

$$\mathfrak{h}(A\ ,\ B) \quad = \quad [1 - \mathrm{F}(A, B)]^{1/2} \quad = \quad \left[1 - \left\|\sqrt{A}\sqrt{B}\right\|_{\mathrm{tr}}\right]^{1/2} \ .$$

This metric generalizes Hellinger distance to quantum states; when A, B are random variables, Bures distance coincides with Hellinger distance. For pure states $|\psi_1\rangle, |\psi_2\rangle$ we use $\mathfrak{h}(|\psi_1\rangle\ ,\ |\psi_2\rangle)$ as shorthand for $\mathfrak{h}(|\psi_1\rangle\langle\psi_1|\ ,\ |\psi_2\rangle\langle\psi_2|)$. Bures distance is related to trace distance in the following manner (see, e.g., Lemma II.6 in Ref. [30]):

Proposition 4.1. *Let P, Q be quantum states over the same Hilbert space. Then*

$$\mathfrak{h}(P\ ,\ Q)^2 \quad \leq \quad \frac{1}{2}\|P - Q\|_{\mathrm{tr}} \quad \leq \quad \sqrt{2}\,\mathfrak{h}(P\ ,\ Q) \ .$$

In the following, let $(p_x), (q_y)$ be distributions over the finite sample spaces $\mathcal{S}, \mathcal{S}'$, respectively.

The Bures distance satisfies the following property.

Proposition 4.2. *Let P_x, Q_x be quantum states over the same finite Hilbert space for each $x \in \mathcal{S}$. Let $P = \sum_{x\in\mathcal{S}} p_x|x\rangle\langle x| \otimes P_x$, and $Q = \sum_{x\in\mathcal{S}} p_x|x\rangle\langle x| \otimes Q_x$. Then*

$$\mathfrak{h}(P\ ,\ Q)^2 \quad = \quad \sum_{x\in\mathcal{S}} p_x\,\mathfrak{h}(P_x\ ,\ Q_x)^2 \ .$$

This may be verified readily by the definition of the Bures distance, but may also be derived as an immediate consequence of the strong concavity property of fidelity [42, Theorem 9.7, p. 414].

The Local Transition Theorem due to Uhlmann [42] helps us find purifications of quantum states that achieve the Bures distance between them.

Proposition 4.3 (Local Transition Theorem)**.** *Let $|\psi_1\rangle$ and $|\psi_2\rangle$ be two pure states in a tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$ of Hilbert spaces. Then there exists a unitary operator U on $\mathcal{H}_1$ such that*

$$\mathfrak{h}((U \otimes \mathbb{I}_{\mathcal{H}_2})\,|\psi_1\rangle\ ,\ |\psi_2\rangle) \quad = \quad \mathfrak{h}(\mathrm{Tr}_{\mathcal{H}_1}|\psi_1\rangle\langle\psi_1|\ ,\ \mathrm{Tr}_{\mathcal{H}_1}|\psi_2\rangle\langle\psi_2|) \ .$$

We rely on a number of standard results from quantum information theory in this work. For a comprehensive introduction to the subject, we refer the reader to a text such as [42].

Let $\mathrm{S}(P)$ denote the von Neumann entropy of the quantum state P, and $\mathrm{I}(P : Q)$ denote the mutual information between the two parts of a joint quantum state PQ.

For a joint quantum state $XQ = \sum_{x\in\mathcal{S}} p_x|x\rangle\langle x| \otimes Q_x$ we define the conditional von Neumann entropy as $\mathrm{S}(Q\,|\,X) = \sum_{x\in\mathcal{S}} p_x\,\mathrm{S}(Q_x)$. Similarly, for a joint state $XPQ = \sum_{x\in\mathcal{S}} p_x|x\rangle\langle x| \otimes (PQ)_x$, where $(PQ)_x$ is a joint state for each $x \in \mathcal{S}$, we define the conditional mutual information as

$$\mathrm{I}(P:Q\,|\,X) \quad = \quad \mathrm{S}(P\,|\,X) + \mathrm{S}(Q\,|\,X) - \mathrm{S}(PQ\,|\,X) \ .$$

The chain rule for mutual information states:

Proposition 4.4 (Chain rule)**.** *Let $XYQ = \sum_{x\in\mathcal{S},y\in\mathcal{S}'} p_x q_y|xy\rangle\langle xy| \otimes Q_{xy}$ be a joint quantum state. Then*

$$\mathrm{I}(XY:Q) \quad = \quad \mathrm{I}(X:Q) + \mathrm{I}(Y:Q\,|\,X) \ .$$

It follows directly from the identity $\mathrm{S}(XQ) = \mathrm{S}(X) + \mathrm{S}(Q|X)$ for joint states XQ of the form $XQ = \sum_{x\in\mathcal{S}} p_x|x\rangle\langle x| \otimes Q_x$.

The Average Encoding Theorem [30, 23] also holds for quantum states. (In fact, it was first formulated in the context of quantum communication.)

Proposition 4.5 (Average encoding theorem)**.** *Let $XQ = \sum_{x\in\mathcal{S}} p_x|x\rangle\langle x| \otimes Q_x$ be a joint quantum state. Then,*

$$\mathbb{E}_{x\leftarrow X}\ \mathfrak{h}(Q_x\ ,\ Q)^2 \quad \leq \quad \kappa\,\mathrm{I}(X:Q) \ ,$$

where κ is the constant $\frac{\ln 2}{2}$.

4.2 Quantum communication and information cost

We briefly describe the model of two-party quantum communication, *à la* Yao [49]. We only consider protocols with classical inputs and outputs. For the basic elements of quantum computation, we refer the reader to a text such as [42].

Informally, two "players", Alice and Bob, hold some number of qubits. When the protocol starts, Alice holds a classical input represented by a bit string $x \in \mathcal{X}$ and similarly Bob holds $y \in \mathcal{Y}$. The qubits in the workspace of the two parties are initialized to a state $|\Phi\rangle$ that is independent of the inputs x, y, and may be entangled across the parties. The protocol consists of some number $t \geq 1$ of rounds of message exchange, in which the two players "play" alternately. Any party may be the first to play. Suppose it is Alice's turn to play. She applies a unitary operator to her workspace qubits, which depends on her input x and the round. Then, Alice sends some of her workspace qubits to Bob. In the next round, Bob's local computation thus involves some qubits previously in Alice's control. At the end of the t rounds of message exchange, the player to receive the last message, say Bob, observes the qubits in his possession according to a measurement that may depend on his input y. The measurement outcome is considered to be the output of the protocol.

More formally, a two-party quantum communication protocol Π is specified as follows. The protocol uses some N qubits, for some positive integer N, so that the associated state space is $(\mathbb{C}^2)^{\otimes N}$. We view this space as a tensor product space $\mathcal{A} \otimes \mathcal{H}_{\mathsf{A},i} \otimes \mathcal{H}_{\mathsf{B},i} \otimes \mathcal{B}$, for each $i = 0, 1, \ldots, t$, with the initial factorization given by $i = 0$, and the factorization at the end of the jth round given $i = j$. This factorization reflects the ownership of the qubits. The space $\mathcal{A}$ contains Alice's input, $\mathcal{B}$ contains Bob's input, and the spaces $\mathcal{H}_{\mathsf{A},i}$ and $\mathcal{H}_{\mathsf{B},i}$ correspond to Alice's and Bob's workspace qubits at the end of round i, respectively.

The qubits in space $\mathcal{A}$ are initialized to $|x\rangle$, and those in $\mathcal{B}$ are initialized to $|y\rangle$. The qubits in the space $\mathcal{H}_{\mathsf{A},0} \otimes \mathcal{H}_{\mathsf{B},0}$ are initialized to a possibly entangled state $|\Phi\rangle$ that is independent of the inputs. The initial joint state is thus $|x\rangle \otimes |\Phi\rangle \otimes |y\rangle$.

The protocol specifies the number t of messages sent, and the player that sends the first message. Suppose it is Alice's turn to play in round i, with $i \geq 1$. The workspace of the two players just before the round factors as $\mathcal{H}_{\mathsf{A},i-1} \otimes \mathcal{H}_{\mathsf{B},i-1}$. Alice applies a unitary operator $V_{i,x}$ to the qubits in $\mathcal{H}_{\mathsf{A},i-1}$. Note that her operator depends on her input x and the round. (Later, we imagine running the protocol on superpositions of inputs. In this case, we think of Alice as applying the unitary $V_i = \sum_x |x\rangle\langle x| \otimes V_{i,x}$ to the qubits in the space $\mathcal{A} \otimes \mathcal{H}_{\mathsf{A},i-1}$.) Then, Alice

sends some of her qubits, corresponding to the space $\mathcal{M}_i$, to Bob. That is, the space $\mathcal{H}_{\mathsf{A},i-1}$ factors as $\mathcal{H}_{\mathsf{A},i} \otimes \mathcal{M}_i$, and $\mathcal{H}_{\mathsf{B},i} = \mathcal{M}_i \otimes \mathcal{H}_{\mathsf{B},i-1}$.

After the tth message is sent, the recipient, say Bob, observes the qubits corresponding to $\mathcal{H}_{\mathsf{B},t}$ according to a POVM (positive operator valued measurement) that depends on his input y. The output of the protocol is the measurement outcome, and we denote the corresponding random variable by $\Pi(x, y)$. Figure 2 depicts such a two-party protocol.

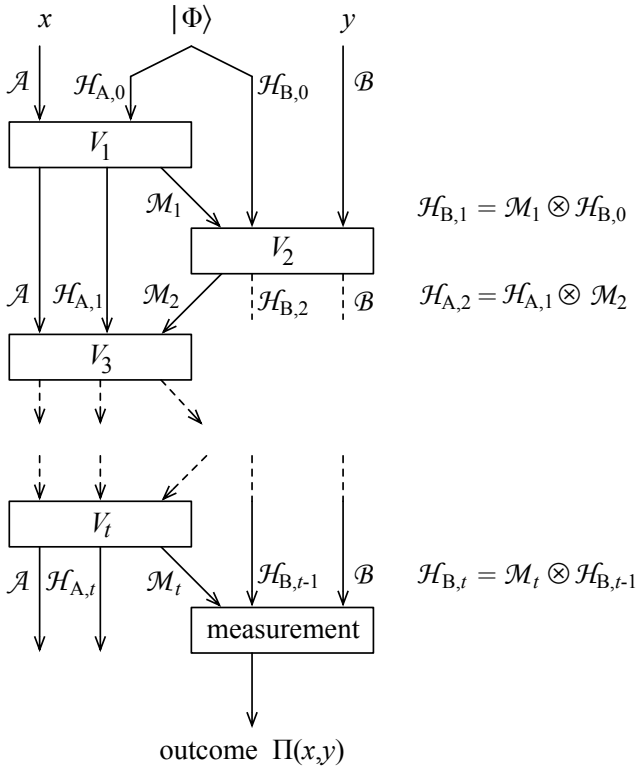

Figure 2: A quantum two-party communication protocol with t messages, inputs x, y and shared initial state $|\Phi\rangle$.

We emphasize that the input qubits in the protocol are *read only*, and that there are no intermediate measurements. A more general protocol may be transformed into this form by appealing to standard techniques in quantum computation [7].

In this article, we are concerned with protocols designed to compute a bi-variate Boolean function $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$. As for classical protocols, the probability of correctness (or *success*) of a protocol on input x, y is $\Pr[\Pi(x, y) = f(x, y)]$. We consider inputs drawn from a joint distribution XY, in which case the success probability is $\Pr[\Pi(X, Y) = f(X, Y)]$. The probability of the complementary event is called the *error* of the protocol on the distribution XY.

As in the classical case, there is no canonical measure of quantum information leaked by a protocol, and this notion is a topic of active research. The choice of the measure is driven by a motivating application and the ease with which we can analyze it. We typically strike a balance between these opposing forces.

A significant difference between classical and quantum information costs arises because the no cloning principle [42, p. 532] prevents the two parties from keeping a copy of the messages. A natural notion of a transcript that encapsulates the history of a quantum protocol is instead the sequence of the joint states after each message exchange. Correspondingly, the notion of information cost is also different from the one in the classical case.

Consider a quantum communication protocol Π with a total of t messages, beginning with

Alice and alternating with Bob. We emphasize that the input qubits in Π are read-only. The first player is assumed to be Alice solely to eliminate awkwardness in defining and referring to quantum information cost. The assumption may be removed without affecting the results in this article. Alternatively, if Bob starts, we may modify the protocol so that Alice sends a single qubit in a fixed state, say $|0\rangle$, at the beginning. This does not affect the information cost, but increases the number of messages by one.

Let λ be a probability distribution over $\mathcal{X} \times \mathcal{Y}$, and let random variables XY be distributed according to λ. Let $P_i Q_i$ denote the joint state of Alice and Bob's workspace *immediately after* the ith message is sent, in a protocol Π when we start with the inputs XY. In analogy with the classical case, we may define the quantum information cost of Π for Alice with respect to λ as

$$\sum_{\text{odd } i \in [t]} \mathrm{I}(X : Q_i \,|\, Y) \ , \tag{4.1}$$

and similarly for Bob as

$$\sum_{\text{even } i \in [t]} \mathrm{I}(Y : P_i \,|\, X) \ . \tag{4.2}$$

A similar definition has been considered by Jain, Radhakrishnan, and Sen [23]. This appears to be a natural definition; it captures the amount of information about the other party's input that is not already contained in her state. It also allows us to relate quantum streaming algorithms for Dyck(2) that use small space, to two-party protocols for Augmented Index with small quantum information cost. (The reduction described in Section 3 extends to quantum algorithms with minor modifications.) However, we are not able to prove an information cost trade-off for Augmented Index with this definition.

The tension between applicability and ease of analysis is rather acute in our case. This leads us to consider the information contained in the messages when the input qubits are initialized to an appropriate *superposition*. This information is in general more than that contained in the messages when we have the corresponding *distribution* over inputs. The former measure may sometimes capture the information revealed by a party in a quantum communication protocol more accurately (see, e.g., Ref. [24]). The resulting notion also seems to be necessary for the proof of the information cost trade-off we present.

Defining quantum information cost with superpositions over inputs, corresponding to arbitrary non-product distributions, comes with its own set of complications. A comprehensive discussion of such measures is beyond the scope of this article. We focus on distributions λ over the input space $\mathcal{X} \times \mathcal{Y}$ with $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2$, and the following limited type of dependence. Let X, Y_1 be independent random variables taking values in $\mathcal{X}, \mathcal{Y}_1$, respectively, and $Y_2 = s(X, Y_1) \in \mathcal{Y}_2$, where s is some function of the first two random variables. Moreover, the function s is such that the conditional random variables $X|(Y_2 = v)$ and $Y_1|(Y_2 = v)$ are also independent, for any v with $\Pr[Y_2 = v] \neq 0$. Then λ is the distribution of XY_1Y_2. In other words, Alice is given some input X, Bob an independent input Y_1, and also a joint function $Y_2 = s(X, Y_1)$ of the two. Moreover, their inputs X, Y_1 remain independent when conditioned on any given value of Y_2. Such distributions include product distributions as well as distributions for problems in which the two communicating parties may share a portion of the input, as in the case of Augmented Index. (The correspondence for Augmented Index is that X is uniformly distributed over $\{0,1\}^n$, Y_1 is the index K that is uniformly distributed over $[n]$, and $Y_2 = s(X, Y_1) = X[1, K]$.)

The final point of difference between the notions of classical and quantum information cost we consider comes from the dependence described above in the distribution λ. Recall that under this distribution λ, Bob's input Y_1 is independent of X and that Bob additionally gets $Y_2 = s(X, Y_1)$. In the classical case, Alice may have information about Y_2 due to its dependence on X, but does not have any information about Y_1, i.e., $\mathrm{I}(X : Y_1) = 0$. When the input registers are initialized with a superposition corresponding to λ, however, Alice may *gain* information about Bob's input Y_1 without any communication between the parties: we may have $\mathrm{I}(\hat{X} : \hat{Y}_1) > 0$, where $\hat{X}\hat{Y}_1\hat{Y}_2$ are in state $\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \sqrt{\lambda(x, y)}\,|x, y\rangle$.

To illustrate this phenomenon, consider the following example. Let X be uniformly distributed over $\{0,1\}^n$, Y_1 be an index K that is uniformly distributed over $[n]$, and $Y_2 = X_K$, i.e., the Kth bit of X. We have $\mathrm{I}(X : Y_1) = 0$. Let $\hat{X}\hat{Y}_1\hat{Y}_2$ be initialized to the state

$$\frac{1}{\sqrt{n2^n}} \sum_{x\in\{0,1\}^n, k\in[n]} |x, k, x_k\rangle \ .$$

Suppose we measure the qubits holding $\hat{Y}_1$ in the basis $(|i\rangle)_{i\in[n]}$ and recover Y_1. By monotonicity of mutual information under quantum operations [42, Theorem 11.15, p. 522], we have $\mathrm{I}(\hat{X} : \hat{Y}_1) \geq \mathrm{I}(\hat{X} : Y_1)$. The reduced state of $\hat{X}Y_1$ is

$$\frac{1}{n} \sum_{k\in[n]} |u\rangle\langle u|^{\otimes(k-1)} \otimes \frac{\mathbb{I}}{2} \otimes |u\rangle\langle u|^{\otimes(n-k)} \otimes |k\rangle\langle k| \ ,$$

where $|u\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. By conjugating $\hat{X}$ by the n-qubit Hadamard operation, we see that the state is equivalent to

$$\frac{1}{n} \sum_{k\in[n]} |0\rangle\langle 0|^{\otimes(k-1)} \otimes \frac{\mathbb{I}}{2} \otimes |0\rangle\langle 0|^{\otimes(n-k)} \otimes |k\rangle\langle k| \ .$$

A straightforward calculation now shows that $\mathrm{I}(\hat{X} : Y_1) = \log_2 n$. So $\mathrm{I}(\hat{X} : \hat{Y}_1) \geq \log_2 n$, whereas $\mathrm{I}(X : Y_1) = 0$.

This phenomenon also occurs in the case of Augmented Index, due to the prefix shared by the two parties. To quantify the information leaked *by the protocol*, rather than *the preparation of the initial state in a superposition*, we view the protocol differently. We imagine that there is a single quantum register that carries the superposition corresponding to X, and that Bob's unitary operations are controlled appropriately by this register. In other words, his transformation in the ith round is of the form

$$V_i \quad = \quad \sum_{x,y_1} |x\rangle\langle x| \otimes |y_1\rangle\langle y_1| \otimes V_{i,y_1 s(x,y_1)} \ ,$$

where the qubits holding x are with Alice. Bob's information cost is then measured with respect to all the qubits with Alice.

We are now in a position to define the measure of quantum information cost for two-party protocols that we analyze. Let λ be a probability distribution over $\mathcal{X} \times \mathcal{Y}$ of the type described above, and let $\hat{X}\hat{Y}_1$ denote the corresponding superposition $\sum_{x\in\mathcal{X}, y_1\in\mathcal{Y}_1} \sqrt{\lambda(x, y_1)}\,|x, y_1\rangle$ over inputs. Let $\hat{X}P_iQ_i\hat{Y}_1$ denote the joint state of Alice and Bob's input and workspace qubits *immediately after* the ith message is sent, in a protocol Π when we start with the input qubits in state $\hat{X}\hat{Y}_1$. Note that the input qubits may get entangled with the message qubits during the protocol. As the state of the input qubits we refer to will be clear from the context, we do not label it with the message number i. The quantum information cost of Π for Bob with respect to λ is then defined as

$$\mathrm{QIC}^{\mathsf{B}}_{\lambda}(\Pi) \quad = \quad \sum_{\text{even } i\in[t]} \mathrm{I}(\hat{Y}_1 : \hat{X}P_i) \ .$$

In this cost, we measure the information about $\hat{Y}_1$ contained in Alice's quantum state, while disregarding $Y_2 = s(X, Y_1)$ (which is not available to Alice).

In Alice's cost, we would like to measure the information about $\hat{X}$ in Bob's quantum state, given access to Y_2. We model this as follows. We imagine an additional register that we label Y_2. We copy $s(X, Y_1)$ into this register and measure the qubits in the standard basis. The initial state of the registers $\hat{X}\hat{Y}_1Y_2$ is then

$$\sum_{y_2\in\mathcal{Y}_2} \sum_{\substack{x,x'\in\mathcal{X} \\ y_1,y_1'\in\mathcal{Y}_1: \\ s(x,y_1)=s(x',y_1')=y_2}} \sqrt{\lambda(x, y_1)\,\lambda(x', y_1')}\ |x, y_1\rangle\langle x', y_1'| \otimes |y_2\rangle\langle y_2| \ .$$

The joint state $\hat{X}P_iQ_i\hat{Y}_1Y_2$ of Alice and Bob's input and workspace qubits, immediately after the ith message is sent, is correspondingly affected. We define Alice's information cost as

$$\mathrm{QIC}^{\mathsf{A}}_{\lambda}(\Pi) \quad = \quad \sum_{\text{odd } i\in[t]} \mathrm{I}(\hat{X} : Q_i\hat{Y}_1 \,|\, Y_2) \ .$$

The inclusion of the register holding Y_2 precisely captures the distribution of inputs in the communication protocol. The artificial construct described before, of substituting this with suitable read-only access to Alice's input qubits (for executing Bob's unitary transformations), however, is more appropriate for the proof of the quantum information cost trade-off.

The above notion corresponds to a hybrid of "internal" and "external information cost" [6]. For product distributions (when Y_2 is trivial), each term of this notion reduces precisely to the amount of (quantum) information available to a party about the other's input.

In the rest of Section 4, we use a convention similar to the one above: a symbol such as Z without a hat denotes the random variable resulting from an imagined measurement, in the computational basis, of a sequence of qubits initialized to a superposition. The state of the qubits prior to the measurement is denoted by the symbol with a hat, e.g., $\hat{Z}$.

Measuring any part of a quantum system in general affects the state of the remaining qubits. Thus the symbol $\hat{X}$ used in the expressions for Alice's and Bob's information cost denotes potentially different states. In the analysis that we present for Augmented Index, we imagine measurements only of parts of Alice's and Bob's inputs in the computational basis. In that case, we denote the resulting state of the qubits without a hat. Thus the state we mean will be clear from the context.

4.3 The quantum information cost trade-off

In this section, we derive an analogue of the information trade-off result established in Section 2.3 for quantum communication protocols for Augmented Index.

We first specialize the notion of quantum information cost to the Augmented Index function f_n, and simplify it further. This allows us derive a stronger information cost trade-off than with the original definition. Let (X, K, B) be random variables distributed according to μ, the uniform distribution over $\{0,1\}^n \times [n] \times \{0,1\}$. Let μ_0 denote the distribution μ conditioned upon $X_K = B$, i.e., when the inputs are chosen uniformly from the set of 0s of f_n. We are interested in the quantum information cost of a protocol Π for Augmented Index under the distribution μ_0, for the two parties.

As explained in Section 4.2, we adopt the following convention with respect to the inputs for Augmented Index. Alice is given the input x. We imagine that Bob is given k, b, and *access* to the prefix $x[1, k-1]$, rather than a copy of these bits. When we restrict to the distribution μ_0, we assume he has read-only access to $x[1, k]$. This means that in any round i of the protocol in which Bob plays, his local unitary operation V_i is controlled by the qubits with Alice that hold the prefix. It is important to bear in mind the qubits on which the unitary operations of the protocol act non-trivially, i.e., do not equal the identity. In particular, in Lemma 4.10, we use the commutativity of the unitary operations used in the protocol and the corresponding unitary operations given by Lemmata 4.8 and 4.9. See, for example, the paragraph before Eq. (4.11).

Suppose we have a quantum protocol Π for Augmented Index with a total of t messages. Without loss of generality (see Section 4.2), we assume that Alice sends the first message, and alternates with Bob thereafter.

Let $\hat{X}P_iQ_i\hat{K}\hat{B}$ denote the joint state of Alice and Bob's workspace in the protocol Π immediately after the ith message is sent, when we start with uniform superpositions $\hat{X}$ over strings $x \in \{0,1\}^n$, $\hat{K}$ over $[n]$, and $\hat{B}$ over $\{0,1\}$ (this corresponds to distribution μ). Let $\hat{X}^0P_i^0Q_i^0\hat{K}^0\hat{B}^0$ denote the analogous joint state corresponding to μ_0, where we assume that Bob is given read-only access to the register containing x_k, rather than a copy of this bit. The quantum information

cost of Π for Alice and Bob with respect to μ_0 is then

$$\begin{aligned} \mathrm{QIC}^{\mathsf{A}}_{\mu_0}(\Pi) &= \sum_{\text{odd } i\in[t]} \mathrm{I}(\hat{X}^0[K+1,n] : Q_i^0\hat{K}^0 \,|\, X[1,K]) \ , \qquad \text{and} \\ \mathrm{QIC}^{\mathsf{B}}_{\mu_0}(\Pi) &= \sum_{\text{even } i\in[t]} \mathrm{I}(\hat{K}^0 : \hat{X}^0 P_i^0) \ . \end{aligned}$$

Due to the monotonicity of mutual information under quantum operations [42, Theorem 11.15, p. 522], for each $i = 1, \ldots, t$ we have

$$\begin{aligned} \mathrm{I}(X : Q_i^0 \,|\, X[1,K]) &\leq \mathrm{I}(\hat{X}^0[K+1,n] : Q_i^0\hat{K}^0 \,|\, X[1,K]) \ , \qquad \text{and} \\ \mathrm{I}(K : \hat{X}^0 P_i^0) &\leq \mathrm{I}(\hat{K}^0 : \hat{X}^0 P_i^0) \ , \end{aligned}$$

where the symbols without a hat denote random variables resulting from an imagined measurement of the corresponding qubits in the computational basis. (We drop the superscript '0' on these random variables, as their marginals are the same as under the distribution μ.) The trade-off we prove also holds for the potentially smaller quantities on the left side above. In order to state the theorem in the strongest possible terms, we define another measure of information cost as follows:

$$\begin{aligned} \tilde{\mathrm{Q}}\mathrm{IC}^{\mathsf{A}}_{\mu_0}(\Pi) &= \sum_{\text{odd } i\in[t]} \mathrm{I}(X : Q_i^0 \,|\, X[1,K]) \ , \qquad \text{and} \\ \tilde{\mathrm{Q}}\mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi) &= \sum_{\text{even } i\in[t]} \mathrm{I}(K : \hat{X}^0 P_i^0) \ . \end{aligned}$$

The intuition behind the lower bound on quantum information cost is the same as that in the classical case. Namely, starting from an input pair on which the function evaluates to 0, if the information cost of any one party is low and we carefully change her input, the other party's share of the state does not change much. Assume for simplicity that Alice produces the output of the protocol. We show that even when we simultaneously change both parts of the input, resulting in a 1-input of the function, the perturbation to Alice's final state is also correspondingly small. This implies that the two information costs cannot be small simultaneously. For more intuition into the main lemmata in this proof, we refer the reader to the analogous steps in the classical case. In the final piece of the argument for the quantum case, the Local Transition Theorem and a hybrid argument take the place of the Cut-and-Paste Lemma. Unlike the latter, these are applied on a message-by-message basis, *à la* Jain, Radhakrishnan, and Sen [23], and leads to a dependence of the information cost trade-off on the number of messages in the protocol.

The next theorem executes this argument for even n. A similar result also holds for odd n, and may be inferred from the proof for the even case. As explained in the previous section, the assumption that Alice sends the first message is not necessary.

Theorem 4.6. *Let Π be any quantum two-party communication protocol for the* Augmented Index *function f_n with n even, Alice starting and alternating with Bob for a total of $t \geq 1$ messages. If Π makes error at most $\varepsilon \in [0, 1/4]$ on the uniform distribution μ over inputs, then*

$$2\left[\frac{\tilde{\mathrm{Q}}\mathrm{IC}^{\mathsf{A}}_{\mu_0}(\Pi)}{n}\right]^{1/2} + \left[2 \cdot \tilde{\mathrm{Q}}\mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi)\right]^{1/2} \geq \frac{1-4\varepsilon}{4\sqrt{\kappa\, t}} \ ,$$

where μ_0 is the uniform distribution over $f_n^{-1}(0)$.

Proof: Consider a protocol Π as in the statement of the theorem. Let the inputs be given by random variables X, K, B, drawn from the distribution μ, let $d \stackrel{\text{def}}{=} \tilde{\mathrm{Q}}\mathrm{IC}^{\mathsf{A}}_{\mu_0}(\Pi)/n$, and let $c \stackrel{\text{def}}{=} \tilde{\mathrm{Q}}\mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi)$.

Let $\hat{X}P_iQ_iKB$ be the joint state of the registers used in the protocol, when the inputs are initialized with a uniform superposition $\hat{X}$ over $x \in \{0,1\}^n$ and random variables K, B, immediately after the ith message in the protocol. Let $d_i = \frac{1}{n} \mathrm{I}(X : Q_i^0 \mid X[1,K])$ for odd $i \in [t]$, and $c_i = \mathrm{I}(K : \hat{X}^0 P_i^0)$ for even $i \in [t]$. So $d = \sum_{\text{odd } i \in [t]} d_i$ and $c = \sum_{\text{even } i \in [t]} c_i$.

We prove the theorem assuming that Alice computes the output of the protocol, i.e., t is even. The proof when Bob computes the output is similar; we point out the main differences along the way. If t is even, we show that the state XP_t^0 is close in trace distance to the state XP_t^1, where XP_t^1 denotes the reduced state XP_t conditioned on the function value being 1, i.e., when $B = \bar{X}_K$. (Note that X is the classical random variable corresponding to the superposition $\hat{X}$.)

Lemma 4.7. *For even t, $\left\| XP_t^0 - XP_t^1 \right\|_{\mathrm{tr}} \quad \leq \quad 1 + 4\sqrt{\kappa\, t}\left[2\sqrt{d} + \sqrt{2c}\right], \quad$ where $\kappa = \frac{\ln 2}{2}$.*

If t is odd, i.e., Bob computes the output of the protocol, we show the same bound on

$$\left\| Q_t^0\, X[1,K] - Q_t^1\, X[1,K-1]\, \bar{X}_K \right\|_{\mathrm{tr}} \ .$$

Since the protocol identifies the two states XP_t^0 and XP_t^1, with average error ε, and trace distance is monotonic under quantum operations [42, Theorem 9.2, p. 406], we have

$$\left\| XP_t^0 - XP_t^1 \right\|_{\mathrm{tr}} \quad \geq \quad 2(1-2\varepsilon) \ .$$

The theorem follows. ∎

We now prove the core of the theorem, i.e., that if Alice computes the output, her final state for the 0 and 1 inputs are close to each other in distribution.

Proof of Lemma 4.7: When we wish to explicitly write a state, say P_i, as a function of the inputs to Alice and Bob, say x and $x[1,k-1], b$ respectively, we write it as $P_i(x; x[1,k-1], b)$. If $b = x_k$, we write Bob's input as $x[1,k]$.

As before, for any $x \in \{0,1\}^n$ and $i \in [n]$, we let $x^{(i)}$ denote the string that equals x in all coordinates except at the ith. Note that $P_t^1 = P_t(X; X[1,K-1], \bar{X}_K)$ is the same mixed state as $P_t(X^{(K)}; X[1,K])$, since X and $X^{(K)}$ are identically distributed. Thus, our goal is to bound

$$\left\| XP_t(X; X[1,K]) - X^{(K)} P_t(X^{(K)}; X[1,K]) \right\|_{\mathrm{tr}} \ .$$

For reasons similar to those the classical case and new ones arising from our proof (an explanation for which is included below), we consider the trace distance between the first term above with $K \in [n/2]$ and the second term with $K \in [n] - [n/2]$. (Recall that in the classical case, we restricted ourselves to $K \in [n] - [n/2]$ in both terms.) Let J be uniformly and independently distributed in $[n/2]$, and let L be uniformly and independently distributed in $[n] - [n/2]$. Then

$$\begin{aligned}
&\left\| XP_t(X; X[1,K]) - X^{(K)} P_t(X^{(K)}; X[1,K]) \right\|_{\mathrm{tr}} \\
&= \left\| \frac{1}{2}\left(XP_t(X; X[1,J]) + XP_t(X; X[1,L]) \right) \right. \\
&\qquad \left. - \frac{1}{2}\left(X^{(J)} P_t(X^{(J)}; X[1,J]) + X^{(L)} P_t(X^{(L)}; X[1,L]) \right) \right\|_{\mathrm{tr}} \\
&\leq 1 + \frac{1}{2} \left\| XP_t(X; X[1,J]) - X^{(L)} P_t(X^{(L)}; X[1,L]) \right\| \\
&= 1 + \frac{1}{2} \left\| X^{(L)} P_t(X^{(L)}; X[1,J]) - X^{(L)} P_t(X^{(L)}; X[1,L]) \right\| \ , \qquad (4.3)
\end{aligned}$$

where we use the fact that X and $X^{(L)}$ are identically distributed, even given the prefix $X[1,J]$, and that the states $XP_t(X; X[1,J])$ and $X^{(L)} P_t(X^{(L)}; X[1,J])$ are therefore identical. So it

suffices to bound the RHS above. If t is odd, we instead bound

$$\begin{aligned}&\left\|Q_t(X;X[1,K])X[1,K]-Q_t(X^{(K)};X[1,K])X[1,K]\right\|_{\mathrm{tr}}\\ &\quad\leq\quad 1+\frac{1}{2}\left\|Q_t(X;X[1,L])X[1,L]-Q_t(X^{(L)};X[1,L])X[1,L]\right\|_{\mathrm{tr}}\quad .\end{aligned}\tag{4.4}$$

The expression for odd t, Eq. (4.4), is similar to the one we had in the classical case: we focus on the case $K \in [n] - [n/2]$ alone.

For every $j \in [n/2], l \in [n] - [n/2]$ and $z \in \{0,1\}^l$, we consider four runs of the protocol Π. The inputs to Alice and Bob in the four runs are summarized in the table below. Only the first l bits of Alice's input are specified. In all four runs, the last $(n-l)$ input bits of Alice are initialized to a uniform superposition over all $(n-l)$-bit strings. The final column gives the notation for the (pure) state corresponding to the registers $\hat{X}[l+1,n]\,P_iQ_i$, which constitute the last $(n-l)$ inputs bits of Alice, her workspace, and that of Bob, immediately after the ith message has been sent, $i \in [t]$.

Run	Alice's input $x[1,l]$	Bob's input $k, x[1,k-1], b$	State
00	z	$j, z[1,j-1], z_j$	$\|\phi_i(z,j)\rangle$
01	z	$l, z[1,l-1], z_l$	$\|\phi_i(z,l)\rangle$
10	$z^{(l)}$	$j, z[1,j-1], z_j$	$\|\phi_i(z^{(l)},j)\rangle$
11	$z^{(l)}$	$l, z[1,l-1], z_l$	$\|\phi_i(z^{(l)},l)\rangle$

The two bits in the "Run" column indicate whether Alice's lth bit has been flipped, and whether we have switched j to l. A "1" indicates a switch. Note that for the first three kinds of inputs, the function value is 0, and for the last it is 1.

When Bob's information cost is low, it follows that the final state on inputs of type "00" is close to the final state on inputs of type "01" (Lemma 4.8). We show a similar closeness between the final state on inputs of type "10" and that on inputs of type "11". This explains the choice made in Eq. (4.3) when Alice produces the output of the protocol. For similar reasons, when Bob produces the output of the protocol, we compare the final state of the protocol on inputs of type "01" with that on inputs of type "11", as in Eq. (4.4).

As the first step, we compare the intermediate protocol states in the above four runs, when we flip the lth input bit of Alice, and when we switch Bob's input from j to l (along with the corresponding prefix). We show that the switch results in a perturbation to reduced state of the other party that is related to the information contained about the bit or the index (as in the classical case). To quantify this perturbation, define

$$h_i(j,l,z) \quad = \quad \mathfrak{h}\Big(Q_i(zX[l+1,n];z[1,j])\ ,\ Q_i(z^{(l)}X[l+1,n];z[1,j])\Big)\quad ,$$

for every odd $i \in [t]$. This is the perturbation in Bob's reduced state when we flip the lth bit of Alice input, when Bob has index j. Define

$$h_i(j,l,z) \quad = \quad \mathfrak{h}\Big(\hat{X}[l+1,n]\,P_i(z\hat{X}[l+1,n];z[1,j])\ ,\ \hat{X}[l+1,n]\,P_i(z\hat{X}[l+1,n];z[1,l])\Big)\quad ,$$

for every even $i \in [t]$. This is the perturbation in Alice's reduced state when we switch Bob's index from j to l. In the above states, P_i is entangled with the qubits holding $\hat{X}$, and is written as a function of $\hat{X}[l+1,n]$ to emphasize this.

The number of qubits Alice and Bob have during the protocol changes with every message. To maintain simplicity of notation, we denote the identity operator in any round on the register holding $\hat{X}[l+1,n]$ and Alice's workspace qubits by $\mathbb{I}_{\mathsf{A}}$ and the identity operator on Bob's workspace qubits by $\mathbb{I}_{\mathsf{B}}$.

We begin by showing that changing Bob's input alone from j to l while keeping Alice's input fixed at $z\hat{X}[l+1,n]$, does not perturb Alice's reduced state in any round of communication by much, provided the corresponding information cost of Bob is small. By the Local Transition Theorem, we then see that Bob may apply a unitary operation to his qubits alone to bring the protocol states close to each other.

Lemma 4.8. *For every even $i \in [t]$, there is a unitary operator U_i that depends upon j, l, z, acts on Bob's workspace qubits alone (i.e., on the register holding state Q_i), and is such that*

$$\mathfrak{h}(\,(\mathbb{I}_{\mathsf{A}} \otimes U_i)\,|\phi_i(z,j)\rangle\ ,\ |\phi_i(z,l)\rangle) \quad = \quad h_i(j,l,z)\ .$$

Moreover,

$$\mathbb{E}_{(j',l',z')\leftarrow(J,L,X[1,L])}\ h_i(j',l',z') \quad \leq \quad \sqrt{8\kappa\, c_i}\ .$$

The proof is presented later in this section.

Next, we show that if the information cost of Alice is small, Bob's state Q_i^0 does not carry much information about X, even given a prefix. Therefore, flipping a bit outside the prefix does not perturb Bob's state by much, and there is a unitary operation on Alice's qubits which brings the joint states close to each other.

Lemma 4.9. *For every odd $i \in [t]$, there is a unitary operator U_i that depends upon j, l, z, acts on the qubits holding $\hat{X}[l+1, n]$ and Alice's workspace qubits (the register holding state P_i), and is such that*

$$\mathfrak{h}\Big(\,(U_i \otimes \mathbb{I}_{\mathsf{B}})\,|\phi_i(z,j)\rangle\ ,\ |\phi_i(z^{(l)},j)\rangle\Big) \quad = \quad h_i(j,l,z)\ .$$

Moreover,

$$\mathbb{E}_{(j',l',z')\leftarrow(J,L,X[1,L])}\ h_i(j',l',z') \quad \leq \quad 4\sqrt{\kappa\, d_i}\ .$$

This is proven later in the section.

There is no quantum counterpart to the Cut-and-Paste lemma, so that unlike in the classical case, the above two lemmata are by themselves not sufficient to conclude the theorem. Instead, we combine these with a hybrid argument to show that switching from chosen 0-inputs of Augmented Index of the type "10" (as defined above) to corresponding 1-inputs of type "11" does not affect the final state by "much".

Lemma 4.10. *Let $(U_i)_{i\in[t]}$, be the unitary operators given by Lemmata 4.8 and 4.9. For every odd $r \in [t]$,*

$$\mathfrak{h}\Big((U_r \otimes \mathbb{I}_{\mathsf{B}})\,|\phi_r(z,l)\rangle\ ,\ |\phi_r(z^{(l)},l)\rangle\Big) \quad \leq \quad h_r(j,l,z) + 2\sum_{i=1}^{r-1} h_i(j,l,z)\ .$$

For every even $r \in [t]$,

$$\mathfrak{h}\Big((\mathbb{I}_{\mathsf{A}} \otimes U_r)|\phi_r(z^{(l)},j)\rangle\ ,\ |\phi_r(z^{(l)},l)\rangle\Big) \quad \leq \quad h_r(j,l,z) + 2\sum_{i=1}^{r-1} h_i(j,l,z)\ .$$

This is proved later in this section.

Recall that t is even. We have

$$\begin{aligned}
&\left\| X^{(L)} P_t(X^{(L)}; X[1,J]) - X^{(L)} P_t(X^{(L)}; X[1,L]) \right\|_{\rm tr} \\
&\leq \mathbb{E}_{(j,l,z)\leftarrow(J,L,X[1,L])} \left\| X[l+1,n]\, P_t(z^{(l)} X[l+1,n]; z[1,j]) - X[l+1,n]\, P_t(z^{(l)} X[l+1,n]; z[1,l]) \right\|_{\rm tr} \\
&\qquad \text{(by the Triangle Inequality)} \\
&\leq \mathbb{E}_{(j,l,z)\leftarrow(J,L,X[1,L])} \left\| \hat{X}[l+1,n]\, P_t(z^{(l)} \hat{X}[l+1,n]; z[1,j]) - \hat{X}[l+1,n]\, P_t(z^{(l)} \hat{X}[l+1,n]; z[1,l]) \right\|_{\rm tr} \\
&\qquad \text{(by the monotonicity of trace distance under quantum operations [42, Theorem 9.2, p. 406])} \\
&\leq 2\sqrt{2}\ \mathbb{E}_{(j,l,z)\leftarrow(J,L,X[1,L])}\ \mathfrak{h}\Big(\hat{X}[l+1,n]\, P_t(z^{(l)} \hat{X}[l+1,n]; z[1,j])\ ,\ \hat{X}[l+1,n]\, P_t(z^{(l)} \hat{X}[l+1,n]; z[1,l]) \Big) \\
&\qquad \text{(by Proposition 4.1)} \\
&\leq 2\sqrt{2}\ \mathbb{E}_{(j,l,z)\leftarrow(J,L,X[1,L])}\ \mathfrak{h}\Big((\mathbb{I}_{\mathsf{A}} \otimes U_t) |\phi_t(z^{(l)}, j)\rangle\ ,\ |\phi_t(z^{(l)}, l)\rangle \Big) \\
&\qquad \text{(by monotonicity of Bures distance under quantum operations [42, Theorem 9.6, p. 414])} \\
&\leq 4\sqrt{2}\ \mathbb{E}_{(j,l,z)\leftarrow(J,L,X[1,L])} \sum_{i=1}^{t} h_i(j,l,z) \qquad \text{(by Lemma 4.10)} \\
&\leq 4\sqrt{2} \left[\sum_{\text{odd } i\in[t]} 4\sqrt{\kappa\, d_i} + \sum_{\text{even } i\in[t]} 2\sqrt{2\kappa\, c_i} \right] \qquad \text{(by Lemmata 4.8 and 4.9)} \\
&\leq 8\sqrt{\kappa\, t} \left[2\sqrt{d} + \sqrt{2c} \right] \ . \qquad \text{(by the Jensen Inequality)}
\end{aligned}$$

In deriving the fourth inequality above, we used the fact that the states here are purification of the states in the previous inequality. This gives us a bound on the RHS of Eq. (4.3), and concludes the proof of Lemma 4.7. ■

We turn to the deferred proofs.

Lemma 4.8. *For every even $i \in [t]$, there is a unitary operator U_i that depends upon j, l, z, acts on Bob's workspace qubits alone (i.e., on the register holding state Q_i), and is such that*

$$\mathfrak{h}(\, (\mathbb{I}_{\mathsf{A}} \otimes U_i)\, |\phi_i(z,j)\rangle\ ,\ |\phi_i(z,l)\rangle) \quad = \quad h_i(j,l,z)\ .$$

Moreover,

$$\mathbb{E}_{(j',l',z')\leftarrow(J,L,X[1,L])}\ h_i(j',l',z') \quad \leq \quad \sqrt{8\kappa\, c_i}\ .$$

Proof: Note that $\hat{X}[l+1,n]\, P_i(z\hat{X}[l+1,n]; z[1,k])$ for $k \leq l$ is the reduced state of $|\phi(z,k)\rangle$ with Bob's workspace (i.e., the register holding state Q_i) traced out. By the Local Transition Theorem, Proposition 4.3, there is a unitary operator U_i that depends upon j, l, z, acts on Bob's workspace qubits alone, and is such that

$$\mathfrak{h}(\, (\mathbb{I}_{\mathsf{A}} \otimes U_i)\, |\phi_i(z,j)\rangle\ ,\ |\phi_i(z,l)\rangle) \quad = \quad h_i(j,l,z)\ .$$

We show that this distance is bounded on average. Consider the quantum state $\hat{X}\tilde{P}_i$ which is the reduced state of all quantum registers except Bob's workspace and his input K. We denote by $\hat{X} P_i(\hat{X}; \hat{X}[1,k])$ this state for a fixed index k, so that

$$\hat{X}\tilde{P}_i \quad = \quad \frac{1}{n} \sum_{k=1}^{n} \hat{X} P_i(\hat{X}; \hat{X}[1,k])\ .$$

By the Average Encoding Theorem, Proposition 4.5,

$$\mathbb{E}_{k\leftarrow K}\ \mathfrak{h}\Big(\hat{X} P_i(\hat{X}\,;\, \hat{X}[1,k])\ ,\ \hat{X}\tilde{P}_i \Big)^2 \quad \leq \quad \kappa\, c_i\ ,$$

where $\kappa = \frac{\ln 2}{2}$. An immediate consequence is that

$$\begin{aligned}\mathbb{E}_{j' \leftarrow J}\, \mathfrak{h}\Big(\hat{X} P_i(\hat{X}\,;\, \hat{X}[1,j'])\,,\, \hat{X}\tilde{P}_i\Big)^2 &\leq 2\,\kappa\, c_i\ , \qquad \text{and}\\ \mathbb{E}_{l' \leftarrow L}\, \mathfrak{h}\Big(\hat{X} P_i(\hat{X}\,;\, \hat{X}[1,l'])\,,\, \hat{X}\tilde{P}_i\Big)^2 &\leq 2\,\kappa\, c_i\ .\end{aligned}$$

By the Triangle Inequality, for any $j' \in [n/2]$, $l' \in [n] - [n/2]$,

$$\begin{aligned}&\mathfrak{h}\Big(\hat{X} P_i(\hat{X}\,;\, \hat{X}[1,j'])\,,\, \hat{X} P_i(\hat{X}\,;\, \hat{X}[1,l'])\Big)^2\\ &\leq \Big(\mathfrak{h}\Big(\hat{X} P_i(\hat{X}\,;\, \hat{X}[1,j'])\,,\, \hat{X}\tilde{P}_i\Big) + \mathfrak{h}\Big(\hat{X} P_i(\hat{X}\,;\, \hat{X}[1,l'])\,,\, \hat{X}\tilde{P}_i\Big)\Big)^2\\ &\leq 2\,\mathfrak{h}\Big(\hat{X} P_i(\hat{X}\,;\, \hat{X}[1,j'])\,,\, \hat{X}\tilde{P}_i\Big)^2 + 2\,\mathfrak{h}\Big(\hat{X} P_i(\hat{X}\,;\, \hat{X}[1,l'])\,,\, \hat{X}\tilde{P}_i\Big)^2\ .\end{aligned}$$

Since Bures distance is monotonic under quantum operations [42, Theorem 9.6, p. 414], measuring the first l' qubits of $\hat{X}$ yields

$$\begin{aligned}&\mathfrak{h}\Big(X[1,l']\,\hat{X}[l'+1,n]\, P_i(X[1,l']\,\hat{X}[l'+1,n]\,;\, X[1,j'])\,,\\ &\qquad X[1,l']\,\hat{X}[l'+1,n]\, P_i(X[1,l']\,\hat{X}[l'+1,n]\,;\, X[1,l'])\Big)^2\\ &\leq 2\,\mathfrak{h}\Big(\hat{X} P_i(X\,;\, X[1,j'])\,,\, \hat{X}\tilde{P}_i\Big)^2 + 2\,\mathfrak{h}\Big(\hat{X} P_i(X\,;\, X[1,l'])\,,\, \hat{X}\tilde{P}_i\Big)^2\ ,\end{aligned}$$

where $X[1,l']$ denotes the classical random variable resulting from the measurement of $\hat{X}[1,l']$. Moreover, by Proposition 4.2, the left hand side above is equal to

$$\mathbb{E}_{z' \leftarrow X[1,l']}\, \mathfrak{h}\Big(\hat{X}[l'+1,n]\, P_i(z'\hat{X}[l'+1,n]\,;\, z'[1,j'])\,,\, \hat{X}[l'+1,n]\, P_i(z'\hat{X}[l'+1,n]\,;\, z'[1,l'])\Big)^2\ .$$

Taking expectation over $(j',l') \leftarrow (J,L)$, and invoking the Jensen inequality, we get the claimed bound. ∎

Lemma 4.9. *For every odd $i \in [t]$, there is a unitary operator U_i that depends upon j,l,z, acts on the qubits holding $\hat{X}[l+1,n]$ and Alice's workspace qubits (the register holding state P_i), and is such that*

$$\mathfrak{h}\Big(\,(U_i \otimes \mathbb{I}_{\mathsf{B}})\,|\phi_i(z,j)\rangle\,,\, |\phi_i(z^{(l)},j)\rangle\Big) \quad = \quad h_i(j,l,z)\ .$$

Moreover,

$$\mathbb{E}_{(j',l',z') \leftarrow (J,L,X[1,L])}\, h_i(j',l',z') \quad \leq \quad 4\sqrt{\kappa\, d_i}\ .$$

Proof: Note that $Q_i(zX[l+1,n]; z[1,k])$ for $k \leq l$ is the reduced state of $|\phi(z,k)\rangle$ with the register holding $\hat{X}$ and Alice's workspace (the register holding state P_i) traced out. By the Local Transition Theorem, Proposition 4.3, there is a unitary operator U_i that depends upon j,l,z, acts on the registers holding $\hat{X}[l+1,n]\, P_i$ alone, and is such that

$$\mathfrak{h}\Big(\,(U_i \otimes \mathbb{I}_{\mathsf{B}})\,|\phi_i(z,j)\rangle\,,\, |\phi_i(z^{(l)},j)\rangle\Big) \quad = \quad h_i(j,l,z)\ .$$

Since $Q_i^0 = Q_i(X\,;\, X[1,K])$, we have

$$\mathrm{I}(X : Q_i(X\,;\, X[1,J])\,|\,X[1,J]) \quad \leq \quad 2\,\mathrm{I}(X : Q_i^0\,|\,X[1,K]) \quad = \quad 2d_i n\ . \tag{4.5}$$

Fix $j' \in [n/2]$ and $z'' \in \{0,1\}^{j'}$. By the Chain Rule, Proposition 4.4,

$$\begin{aligned}&\mathrm{I}(X[j'+1,n] : Q_i(z''X[j'+1,n]\,;\, z''))\\ &= \sum_{l'=j'+1}^{n} \mathrm{I}(X_{l'} : Q_i(z''X[j'+1,n]\,;\, z'')\,|\,X[j'+1,l'-1])\\ &\geq \sum_{l'=n/2+1}^{n} \mathrm{I}(X_{l'} : Q_i(z''X[j'+1,n]\,;\, z'')\,|\,X[j'+1,l'-1])\ .\end{aligned} \tag{4.6}$$

Moreover by the Triangle Inequality, and the Average Encoding Theorem (Proposition 4.5), for any given $l' \in [n] - [n/2]$ and $z' \in \{0,1\}^{l'}$,

$$\begin{aligned}
&\mathfrak{h}\Big(Q_i(z'X[l'+1,n]\,;\,z'[1,j'])\;,\;Q_i(z'^{(l')}X[l'+1,n]\,;\,z'[1,j'])\Big)\\
&\leq\quad \mathfrak{h}(Q_i(z'X[l'+1,n]\,;\,z'[1,j'])\;,\;Q_i(z'[1,l'-1]\,X_{l'}X[l'+1,n]\,;\,z'[1,j']))\\
&\qquad+\mathfrak{h}\Big(Q_i(z'^{(l')}X[l'+1,n]\,;\,z'[1,j'])\;,\;Q_i(z'[1,l'-1]\,X_{l'}X[l'+1,n]\,;\,z'[1,j'])\Big)\\
&\leq\quad [\,4\kappa\,\mathrm{I}(X_{l'}:Q_i(z'[1,l'-1]\,X_{l'}\,X[l'+1,n]\,;\,z'[1,j']))\,]^{1/2}\quad . \qquad (4.7)
\end{aligned}$$

Combining Eqs. (4.5), (4.6), and (4.7), we get

$$\begin{aligned}
&\mathbb{E}_{(j',l',z')\leftarrow(J,L,X[1,L])}\,\mathfrak{h}\Big(Q_i(z'X[l'+1,n]\,;\,z'[1,j'])\;,\;Q_i(z'^{(l')}X[l'+1,n]\,;\,z'[1,j'])\Big)^2\\
&\leq\quad 4\kappa\,\mathbb{E}_{(j',l',z')\leftarrow(J,L,X[1,L])}\,\mathrm{I}(X_{l'}:Q_i(z'[1,l'-1]\,X_{l'}X[l'+1,n]\,;\,z'[1,j']))\\
&=\quad 4\kappa\,\mathbb{E}_{(j',l',z'')\leftarrow(J,L,X[1,J])}\,\mathrm{I}(X_{l'}:Q_i(z''\,X[j'+1,n]\,;\,z'')\,|\,X[j'+1,l'-1])\\
&\leq\quad \frac{8\kappa}{n}\,\mathrm{I}(X:Q_i(X\,;\,X[1,J])\,|\,X[1,J])\quad\leq\quad 16\kappa\,d_i\;\;,
\end{aligned}$$

as claimed. ■

Lemma 4.10. *Let $(U_i)_{i\in[t]}$, be the unitary operators given by Lemmata 4.8 and 4.9. For every odd $r \in [t]$,*

$$\mathfrak{h}\Big((U_r\otimes\mathbb{I}_{\mathsf{B}})\,|\phi_r(z,l)\rangle\;,\;|\phi_r(z^{(l)},l)\rangle\Big)\quad\leq\quad h_r(j,l,z)+2\sum_{i=1}^{r-1}h_i(j,l,z)\;\;.$$

For every even $r \in [t]$,

$$\mathfrak{h}\Big((\mathbb{I}_{\mathsf{A}}\otimes U_r)|\phi_r(z^{(l)},j)\rangle\;,\;|\phi_r(z^{(l)},l)\rangle\Big)\quad\leq\quad h_r(j,l,z)+2\sum_{i=1}^{r-1}h_i(j,l,z)\;\;.$$

Proof: We prove the lemma by induction over $r \in [t]$. The base case is $r = 1$. By the convention we have adopted, Alice sends the first message. Since the joint state immediately after the first message is independent of Bob's input, we have

$$|\phi_1(z,l)\rangle\quad=\quad|\phi_1(z,j)\rangle\qquad\text{and}\qquad|\phi_1(z^{(l)},l)\rangle\quad=\quad|\phi_1(z^{(l)},j)\rangle\;\;.$$

That is, the state on the input of type "01" equals that on the input of type "00". The same holds for inputs of type "11" and "10". Along with Lemma 4.9 we get

$$\begin{aligned}
&\mathfrak{h}\Big((U_1\otimes\mathbb{I}_{\mathsf{B}})\,|\phi_1(z,l)\rangle\;,\;|\phi_1(z^{(l)},l)\rangle\Big)\\
&\quad=\quad\mathfrak{h}\Big((U_1\otimes\mathbb{I}_{\mathsf{B}})\,|\phi_1(z,j)\rangle\;,\;|\phi_1(z^{(l)},j)\rangle\Big)\quad=\quad h_1(j,l,z)\;\;.
\end{aligned}$$

In other words, the state on the input of type "01" is, up to a unitary operation on Alice's part, "close" to that on the input of type "11".

We prove that the lemma holds for r, assuming that it holds for $r-1 \in [t]$. The argument here follows the same intuition as in the base case, but is more involved because the analogous equalities need not hold. However, the first pair of states may be shown to be close to each other, modulo a local unitary operator, by virtue of Bob's low information cost. The second pair are assumed to be close, again modulo a local unitary operator, by the inductive hypothesis. A careful hybrid argument then gives us the claimed bound. Figure 3 depicts this schematically.

There are two cases: r is odd, or r is even. We conduct the argument in the second case, when r is even. The argument for r odd is similar, and is omitted.

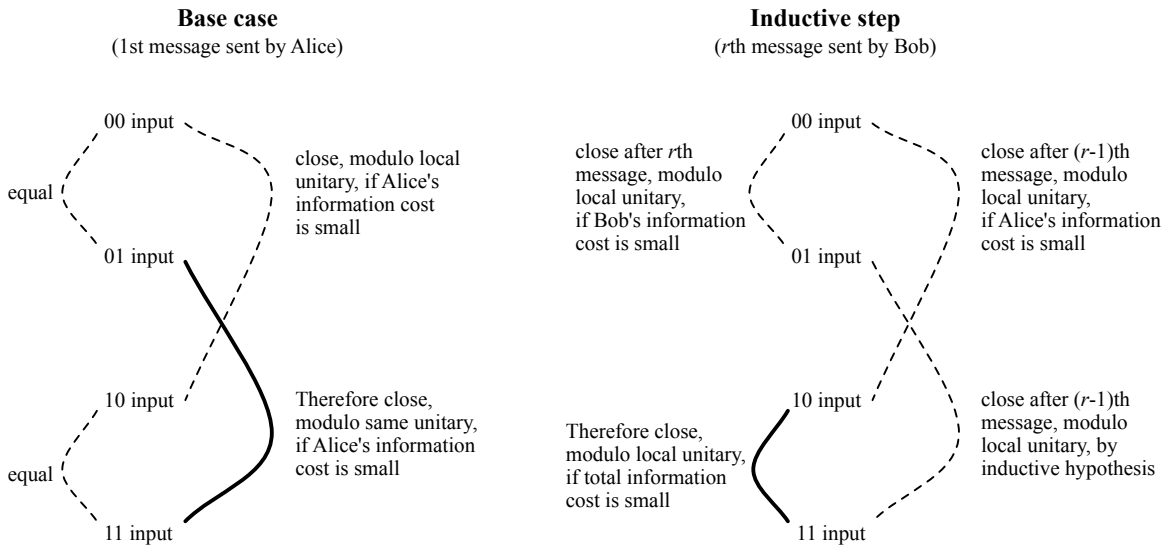

Figure 3: The relationship between states at intermediate stages of the protocol, as described in the proof of Lemma 4.10.

By our convention, Bob sends the even numbered messages, including the rth message. By Lemma 4.8, the states on the inputs of type "00" and "01" are "close" up to the local unitary U_r, i.e.,

$$\mathfrak{h}\left((\mathbb{I}_{\mathsf{A}} \otimes U_r) \left|\phi_r(z,j)\right\rangle \ , \ \left|\phi_r(z,l)\right\rangle \right) \quad = \quad h_r(j,l,z) \ . \tag{4.8}$$

Similarly, by Lemma 4.9, the states *before* the rth message on the inputs of type "00" and "10" are "close" up to the local unitary U_{r-1}, i.e.,

$$\mathfrak{h}\Big((U_{r-1} \otimes \mathbb{I}_{\mathsf{B}}) \left|\phi_{r-1}(z,j)\right\rangle \ , \ \left|\phi_{r-1}(z^{(l)},j)\right\rangle \Big) \quad = \quad h_{r-1}(j,l,z) \ . \tag{4.9}$$

By the induction hypothesis, we also have the following relationship between the states on inputs of type "01" and "11":

$$\mathfrak{h}\Big((U_{r-1} \otimes \mathbb{I}_{\mathsf{B}}) \left|\phi_{r-1}(z,l)\right\rangle \ , \ \left|\phi_{r-1}(z^{(l)},l)\right\rangle \Big) \quad \leq \quad h_{r-1}(j,l,z) + 2\sum_{i=1}^{r-2} h_i(j,l,z) \ . \tag{4.10}$$

Now

$$\begin{aligned} \left|\phi_r(z,l)\right\rangle \quad &= \quad (\mathbb{I}_{\mathsf{A}} \otimes V_{r,z[1,l]}) \left|\phi_{r-1}(z,l)\right\rangle \ , \qquad \text{and} \\ \left|\phi_r(z^{(l)},l)\right\rangle \quad &= \quad (\mathbb{I}_{\mathsf{A}} \otimes V_{r,z[1,l]}) \left|\phi_{r-1}(z^{(l)},l)\right\rangle \ , \end{aligned}$$

where $V_{r,z[1,l]}$ is the unitary operator that Bob applies on his part of the state (i.e., on the register holding state Q_{r-1} before sending the rth message. Note that $V_{r,z[1,l]}$ commutes with U_{r-1}, as they act on disjoint sets of qubits. Since the Bures distance is invariant under unitary operators, Eq. (4.9) gives us

$$\mathfrak{h}\Big((U_{r-1} \otimes \mathbb{I}_{\mathsf{B}}) \left|\phi_r\right\rangle\!(z,j) \ , \ \left|\phi_r(z^{(l)},j)\right\rangle \Big) \quad = \quad h_{r-1}(j,l,z) \ , \tag{4.11}$$

and Eq. (4.10) gives us

$$\mathfrak{h}\Big((U_{r-1} \otimes \mathbb{I}_{\mathsf{B}}) \left|\phi_r(z,l)\right\rangle \ , \ \left|\phi_r(z^{(l)},l)\right\rangle \Big) \quad \leq \quad h_{r-1}(j,l,z) + 2\sum_{i=1}^{r-2} h_i(j,l,z) \ . \tag{4.12}$$

By the Triangle Inequality, Eqs. (4.8), (4.11), and (4.12), and the observation that U_{r-1} and U_r act on disjoint sets of qubits, we get

$$\begin{aligned}
&\mathfrak{h}\Big((\mathbb{I}_{\mathsf{A}} \otimes U_r)\,|\phi_r(z^{(l)}, j)\rangle\ ,\ |\phi_r(z^{(l)}, l)\rangle\Big) \\
&\quad\leq\ \mathfrak{h}\Big((\mathbb{I}_{\mathsf{A}} \otimes U_r)\,|\phi_r(z^{(l)}, j)\rangle\ ,\ (U_{r-1} \otimes \mathbb{I} \otimes U_r)\,|\phi_r(z, j)\rangle\Big) \\
&\qquad + \mathfrak{h}\Big((U_{r-1} \otimes \mathbb{I} \otimes U_r)\,|\phi_r(z, j)\rangle\ ,\ |\phi_r(z^{(l)}, l)\rangle\Big) \\
&\quad=\ h_{r-1}(j, l, z) + \mathfrak{h}\Big((U_{r-1} \otimes \mathbb{I} \otimes U_r)\,|\phi_r(z, j)\rangle\ ,\ |\phi_r(z^{(l)}, l)\rangle\Big) \\
&\quad\leq\ h_{r-1}(j, l, z) + \mathfrak{h}((U_{r-1} \otimes \mathbb{I} \otimes U_r)\,|\phi_r(z, j)\rangle\ ,\ (U_{r-1} \otimes \mathbb{I}_{\mathsf{B}})\,|\phi_r(z, l)\rangle) \\
&\qquad + \mathfrak{h}\Big((U_{r-1} \otimes \mathbb{I}_{\mathsf{B}})\,|\phi_r(z, l)\rangle\ ,\ |\phi_r(z^{(l)}, l)\rangle\Big) \\
&\quad\leq\ h_{r-1}(j, l, z) + h_r(j, l, z) + \mathfrak{h}\Big((U_{r-1} \otimes \mathbb{I}_{\mathsf{B}})\,|\phi_r(z, l)\rangle\ ,\ |\phi_r(z^{(l)}, l)\rangle\Big) \\
&\quad\leq\ h_r(j, l, z) + 2\sum_{i=1}^{r-1} h_i(j, l, z)\ .
\end{aligned}$$

(The identity operators without a subscript in this derivation act on the space of the rth message.) This completes the induction step. ■

5 Concluding remarks

The main focus of this article is the amount of information two parties necessarily reveal about their inputs in the process of the computing a function in a distributed manner. The function of interest is Augmented Index, a natural variant of the Index function that is ubiquitous in communication complexity. We show that in any randomized communication protocol that computes this function correctly with constant error on the uniform distribution (a "hard" distribution), either Alice reveals $\Omega(n)$ information about her n-bit input, or Bob reveals $\Omega(1)$ information about his $(\log n)$-bit input, even when the inputs are drawn from the uniform distribution over inputs which evaluate to 0. At first glance, a trade-off under a distribution on inputs on which the function value is *known in advance* may appear to be counter-intuitive. This is a consequence of the correctness of the protocol on the hard distribution. Such a phenomenon was first demonstrated by Bar-Yossef, Jayram, Kumar, and Sivakumar [5].

The motivation for this work comes from the study of tasks that may be accomplished with a few sequential scans of massive data, using significantly smaller memory, i.e., through streaming algorithms. The above result has implications for the space required by streaming algorithms for Dyck(2), the problem of checking the syntax of a parenthesized expression. It implies that for this problem, we need space $\sqrt{n}/T$ on inputs of length n, when allowed T unidirectional passes over the input.

The proof of the information cost trade-off showcases a modular and conceptually simple technique involving the Average Encoding Theorem and the Cut-and-Paste Lemma. Originally developed to analyse properties of quantum protocols, Average Encoding has been used more widely in classical complexity theory. For instance, it has been used to derive lower bounds for data structures [46], and can be used to derive the "Disguising Distribution Lemma" [17], which has applications for instance compression. The technique developed in this article has also been adapted by François and Magniez to prove space lower bounds for the problem of checking priority queues with time stamps in the streaming model [19]. We expect that these tools have yet more applications in information processing.

A few recent works show how simple *quantum* streaming algorithms may use exponentially smaller amount of space as compared with classical ones [35, 21]. We ask if there is similar advantage in solving a natural and important problem such as Dyck(2). We make partial

progress in this direction, by establishing a *quantum* information cost trade-off for AUGMENTED INDEX. We show that in quantum protocols that compute AUGMENTED INDEX correctly with constant error on the uniform distribution, either Alice reveals $\Omega(n/t)$ information, or Bob reveals $\Omega(1/t)$ information, where t is the number of messages in the protocol, even when the inputs are drawn from the aforementioned easy distribution.

The quantum information cost trade-off by itself does not imply a space lower bound for streaming quantum algorithms. The reduction from streaming algorithms for DYCK(2) with small space to quantum two-party protocols for AUGMENTED INDEX breaks down for the notion of information cost we adopt. We conjecture a trade-off similar to Theorem 4.6 for the notion of information cost in Eqs. (4.1) and (4.2). We leave the resolution of this conjecture as an intriguing open problem.

References

[1] Scott Aaronson. The learnability of quantum states. *Proceedings of the Royal Society A, Mathematical, Physical & Engineering Sciences*, 463(2088):3089–3114, 2007.

[2] Andris Ambainis, Ashwin Nayak, Amnon Ta-Shma, and Umesh Vazirani. Dense quantum coding and a lower bound for 1-way quantum automata. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, pages 376–383. ACM Press, May 1–4, 1999.

[3] Andris Ambainis, Ashwin Nayak, Amnon Ta-Shma, and Umesh Vazirani. Dense quantum coding and quantum finite automata. *Journal of the ACM*, 49(4):1–16, July 2002.

[4] Ziv Bar-Yossef, T. S. Jayram, Robert Krauthgamer, and Ravi Kumar. The sketching complexity of pattern matching. In Klaus Jansen, Sanjeev Khanna, José D. P. Rolim, and Dana Ron, editors, *Proceedings of the 7th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX 2004) and 8th International Workshop on Randomization and Computation (RANDOM 2004)*, volume 3122 of *Lecture Notes in Computer Science*, pages 261–272. Springer, 2004.

[5] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004. Special issue on FOCS 2002.

[6] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. *SIAM Journal on Computing*, 42(3):1327–1363, 2013.

[7] Ethan Bernstein and Umesh V. Vazirani. Quantum complexity theory. *SIAM Journal on Computing*, 26(5):1411–1473, 1997.

[8] Robin Blume-Kohout, Sarah Croke, and Daniel Gottesman. Streaming universal distortion-free entanglement concentration. *IEEE Transactions on Information Theory*, 60(1):334–350, Jan 2014.

[9] Joshua Brody, Harry Buhrman, Michal Koucký, Bruno Loff, Florian Speelman, and Nikolay Vereshchagin. Towards a reverse Newman's theorem in interactive information complexity. In *2013 IEEE Conference on Computational Complexity (CCC)*, pages 24–33, June 2013.

[10] Amit Chakrabarti, Ranganath Kondapally Graham Cormode, and Andrew McGregor. Information cost tradeoffs for Augmented Index and streaming language recognition. Technical Report TR10-076, Electronic Colloquium on Computational Complexity, `http://eccc.hpi-web.de/`, April 18 2010.

[11] Amit Chakrabarti, Ranganath Kondapally Graham Cormode, and Andrew McGregor. Information cost tradeoffs for augmented index and streaming language recognition. *SIAM Journal on Computing*, 42(1):61–83, 2013.

[12] Amit Chakrabarti and Ranganath Kondapally. Everywhere-tight information cost tradeoffs for Augmented Index. In *Proceedings of the 14th International Workshop on Approximation*

Algorithms for Combinatorial Optimization Problems, and the 15th International Workshop on Randomization and Computation, APPROX'11/RANDOM'11, pages 448–459, Berlin, Heidelberg, 2011. Springer-Verlag.

[13] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew C.-C. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 270–278, 2001.

[14] Noam Chomsky and M. P. Schotzenberger. Computer programming and formal languages. In P. Braffort and D. Hirschberg, editors, *The Algebraic Theory of Context-Free Languages*, pages 118–161, Amsterdam, 1963. North Holland.

[15] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley & Sons, New York, NY, USA, 1991.

[16] Khanh Do Ba, Piotr Indyk, Eric Price, and David P. Woodruff. Lower bounds for sparse recovery. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '10, pages 1190–1197, Philadelphia, PA, USA, 2010. Society for Industrial and Applied Mathematics.

[17] Andrew Drucker. New limits to classical and quantum instance compression. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science*, pages 609–618, Los Alamitos, CA, USA, October 20–23 2012. IEEE Computer Society.

[18] Omar Fawzi, Patrick Hayden, and Pranab Sen. From low-distortion norm embeddings to explicit uncertainty relations and efficient information locking. *Journal of the ACM*, 60(6):44:1–44:61, November 2013.

[19] Nathanaël François and Frédéric Magniez. Streaming complexity of checking priority queues. In Natacha Portier and Thomas Wilke, editors, *30th International Symposium on Theoretical Aspects of Computer Science*, volume 20 of *Leibniz International Proceedings in Informatics*, pages 454–465, Dagstuhl, Germany, 2013. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[20] Dmitry Gavinsky and Tsuyoshi Ito. Quantum fingerprints that keep secrets. *Quantum Information and Computation*, 13(7-8):583–606, 2013.

[21] Dmitry Gavinsky, Julia Kempe, Iordanis Kerenidis, Ran Raz, and Ronald de Wolf. Exponential separation for one-way quantum communication complexity, with applications to cryptography. *SIAM Journal on Computing*, 38(5):1695–1708, 2008.

[22] Rahul Jain and Ashwin Nayak. The space complexity of recognizing well-parenthesized expressions. Technical Report TR10-071, Electronic Colloquium on Computational Complexity, `http://eccc.hpi-web.de/`, April 19 2010.

[23] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A lower bound for the bounded round quantum communication complexity of Set Disjointness. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, pages 220–229. IEEE Computer Society Press, Los Alamitos, CA, USA, 2003.

[24] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A property of quantum relative entropy with an application to privacy in quantum communication. *Journal of the ACM*, 56(6):1–32, 2009.

[25] T. S. Jayram, Ravi Kumar, and D. Sivakumar. Two applications of information complexity. In *Proceedings of the Thirty-Fifth annual ACM Symposium on Theory of Computing*, pages 673–682. ACM, 2003.

[26] Daniel M. Kane, Jelani Nelson, and David P. Woodruff. On the exact space complexity of sketching and streaming small norms. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '10, pages 1161–1178, Philadelphia, PA, USA, 2010. Society for Industrial and Applied Mathematics.

[27] Iordanis Kerenidis and Ronald de Wolf. Exponential lower bound for 2-query locally decodable codes. *Journal of Computer and System Sciences*, 69(3):395–420, 2004. Special issue for STOC 2003.

[28] Hartmut Klauck. Quantum and approximate privacy. *Theory of Computing Systems*, 37(1):221–246, 2004.

[29] Hartmut Klauck. One-way communication complexity and the Nečiporuk lower bound on formula size. *SIAM Journal on Computing*, 37(2):552–583, 2007.

[30] Hartmut Klauck, Ashwin Nayak, Amnon Ta-Shma, and David Zuckerman. Interaction in quantum communication. *IEEE Transactions on Information Theory*, 53(6):1970–1982, June 2007.

[31] Ilan Kremer, Noam Nisan, and Dana Ron. On randomized one-round communication complexity. *Computational Complexity*, 8(1):21–49, 1999.

[32] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, Cambridge, UK, 1997.

[33] Lucien Marie Le Cam and Grace Lo Yang. *Asymptotics in Statistics: Some Basic Concepts*. Springer Series in Statistics. Springer-Verlag, New York, 1990.

[34] François Le Gall. Exponential separation of quantum and classical online space complexity. In *Proceedings of the Eighteenth Annual ACM Symposium on Parallelism in Algorithms and Architectures*, SPAA '06, pages 67–73, New York, NY, USA, 2006. ACM.

[35] François Le Gall. Exponential separation of quantum and classical online space complexity. *Theory of Computing Systems*, 45:188–202, 2009.

[36] Richard J. Lipton and Yechezkel Zalcstein. Word problems solvable in logspace. *Journal of the ACM*, 24:522–526, July 1977.

[37] Frédéric Magniez, Claire Mathieu, and Ashwin Nayak. Recognizing well-parenthesized expressions in the streaming model. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing*, pages 261–270, New York, NY, June 6–8 2010. ACM Press.

[38] Peter Bro Miltersen, Noam Nisan, Shmuel Safra, and Avi Wigderson. On data structures and asymmetric communication complexity. *Journal of Computer and System Sciences*, 57(1):37–49, 1998.

[39] S. Muthukrishnan. *Data Streams: Algorithms and Applications*, volume 1, number 2 of *Foundations and Trends in Theoretical Computer Science*. Now Publishers Inc., Hanover, MA, USA, 2005.

[40] Ashwin Nayak. Optimal lower bounds for quantum automata and random access codes. In *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science*, pages 369–376. IEEE Computer Society Press, October 17–19, 1999.

[41] Ilan Newman. Private vs. common random bits in communication complexity. *Information Processing Letters*, 39(2):67–71, 1991.

[42] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, UK, 2000.

[43] Jonathan Oppenheim and Stephanie Wehner. The uncertainty principle determines the nonlocality of quantum mechanics. *Science*, 330(6007):1072–1074, 2010.

[44] Marcin Pawowski, Tomasz Paterek, Dagomir Kaszlikowski, Valerio Scarani, Andreas Winter, and Marek Żukowski. Information causality as a physical principle. *Nature*, 461:1101–1104, 2009.

[45] Michael Saks and Xiaodong Sun. Space lower bounds for distance approximation in the data stream model. In *Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing*, pages 360–369. ACM, 2002.

[46] Pranab Sen and S. Venkatesh. Lower bounds for predecessor searching in the cell probe model. *Journal of Computer and System Sciences*, 74(3):364–385, May 2008.

[47] Thomas Vidick and Stephanie Wehner. Does ignorance of the whole imply ignorance of the parts? Large violations of noncontextuality in quantum theory. *Physical Review Letters*, 107(030402), 2011.

[48] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing. In *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing*, STOC '79, pages 209–213, New York, NY, USA, 1979. ACM.

[49] Andrew Chi-Chih Yao. Quantum circuit complexity. In *Proceedings of the 34th Annual IEEE Symposium on Foundations of Computer Science*, pages 352–361, Los Alamitos, CA, USA, 1993. IEEE Computer Society Press.