Electronic Colloquium on Computational Complexity, Revision 2 of Report No. 71 (2010)

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

Rahul Jain* Ashwin Nayak†

July 20, 2011

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 bi-directional 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. We further demonstrate the effectiveness of these techniques by extending the result to quantum protocols. We show that 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.

*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.

†Department of Combinatorics and Optimization, and Institute for Quantum Computing, University of Waterloo; and Perimeter Institute for Theoretical Physics; 200 University Ave. W., Waterloo, ON, N2L 3G1, Canada. Email: ashwin.nayak@uwaterloo.ca. Work done in part 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.

ISSN 1433-8092

1 Introduction

Streaming algorithms 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. We refer the reader to the text [25] for a more thorough introduction to this topic.

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 study of 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.

Definition 1.1 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 [10], 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 heights, and check parentheses at the same height. This scheme does not seem to easily translate to streaming algorithms, even with a small number of passes over the input. In fact, 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. Another route 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 algorithm for solving the word problem [22] 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.

Theorem 1.1 *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 3.2 later in this article. (Similarly, the theorems we state below are made more precise in later sections.)

DYCK(2) was first studied in the context of the streaming model by Magniez, Mathieu, and Nayak [23], spurred by its practical relevance, e.g., its relationship to the processing of large XML

files, and the connection between formal language theory and complexity in the context of processing massive data. They overcome the apparent difficulties described above and present two 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 establish an optimal space lower bound for one-pass algorithms with polynomially small error. They prove that any one-pass algorithm that makes error at most $1/n \log n$ uses space $\Omega(\sqrt{n \log n})$.

Perhaps more surprisingly, Magniez *et al.* show that the demand on space shrinks drastically when the algorithm is allowed another pass over the input. The second algorithm makes two passes over the input, uses only $O(\log^2 n)$ space, and has polynomially small probability of error. A curious property of the second algorithm is that it makes the second pass in *reverse* order, and this seems essential for its performance. An obvious question then 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. The logarithmic space algorithm for DYCK(2) mentioned above translates to streaming algorithms with a linear number of passes, and suggests the possibility of algorithms with fewer (but more than one) passes, that use sub-polynomial space. Nonetheless, 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 the above conjecture and confirms the intuition that the ability to scan the input in either direction gives streaming algorithms additional computational firepower. The bound we get for one-pass algorithms is a factor of $\sqrt{\log n}$ better than the one in Ref. [23] for constant error probability, but falls short of optimal by the same factor for polynomially small error.

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 this variant, the player holding the index also receives a portion of the other party's input. More formally, 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 $b \in \{0, 1\}$. 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 as "serial encoding" [2, 26], and as "Augmented Index" [12, 17] and "Mountain problem" [23] later works. Informally speaking, we show that in any communication protocol that computes f_n correctly with constant error on the uniform distribution μ (a "hard distribution"), either Alice reveals $\Omega(n)$ information about her input x, or Bob reveals $\Omega(1)$ information about his 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 2.2, 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* $\mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi) \in \Omega(1)$.

The connection between the AUGMENTED INDEX function f_n and streaming algorithms for DYCK(2) was charted by Magniez *et al.* They map a streaming algorithm for DYCK(2) that uses space s to a multi-party communication protocol in which the messages are each of the same length s, and then bound s from below for protocols resulting from one-pass algorithms. The communication bound is derived using the information cost approach (see, for example, Refs. [9, 28, 5, 16, 14]), which reduces the task to bounding from below the information cost of AUGMENTED INDEX.

A notion of information cost for INDEX was studied previously by Jain, Radhakrishnan, and Sen [15]

in the context of privacy in communication. 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 the 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. We note that a stronger trade-off was established by Magniez, Mathieu, and Nayak [23] for two-message protocols that start with Alice, and make polynomially small error. 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 with our techniques, as we do not restrict ourselves to this special form of protocol. Indeed, 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 [7] 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 that they ultimately employ (e.g., the Chain Rule for mutual information and the Pinskert Inequality) are equivalent to ours, they take a different, rather technical, route to these tools. The first version of our article [13] and that of Chakrabarti *et al.* [8] contained trade-offs that were weaker, albeit in different respects. After learning about each other's works, both groups strengthed our respective proofs to achieve qualitatively the same results.

We demonstrate the power of the approach we take by extending it to quantum protocols for AUGMENTED INDEX. Starting with appropriate notions of quantum information cost $(\mathrm{QIC}^{\mathsf{A}}_{\lambda}(\Pi), \mathrm{QIC}^{\mathsf{B}}_{\lambda}(\Pi))$ for a protocol Π for AUGMENTED INDEX, we 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* $\mathrm{QIC}^{\mathsf{B}}_{\mu_0}(\Pi) \in \Omega(1/t)$.

The quantum information cost trade-off involves a number of subtleties, such as quantifying information cost in the absence of a notion of a message transcript, one which also avoids any information leakage due to the non-product nature of the input distribution. The absence of an analogue to the Cut-and-Paste property introduces further complications. We circumvent the Cut-and-Paste property by adapting a hybrid argument due to Jain, Radhakrishnan, and Sen [14] that allows us to analyze quantum protocols one message at a time. These issues are discussed in detail in Section 4.2.

We are not aware of quantum protocols that beat the classical information bounds, and believe the dependence of the trade-off in Theorem 1.3 on the number of rounds t is a consequence of the proof technique. The proof of the connection between quantum streaming algorithms and quantum protocols for AUGMENTED INDEX breaks down in the process of defining a suitable notion of quantum information cost. We leave the possible implications for space lower bounds for quantum streaming algorithms to future work. Finally, we remark that the approaches taken by Magniez *et*

al. [23] and Chakrabarti *et al.* [7] for showing information cost trade-off in classical protocols do not seem to generalize to quantum protocols. They are based on analyzing the input distribution conditioned on the message transcript, for which no suitable quantum analogue is known.

Communication problems involving the INDEX function 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 [21]. For instance, they have been used to analyze data structures [24], the size of finite automata [3] and formulae [19], the length of locally decodable codes [18], learnability of quantum states [1], and sketching complexity [4]. We believe that the more nuanced properties of the INDEX function such as the one we establish here be of fundamental importance, and be likely to find application in other contexts as well.

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. Then we develop the lower bound for classical protocols for AUGMENTED INDEX in Section 2.2.

2.1 Information theory and communication complexity 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 often 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 $i, j \in [n] = \{1, 2, \ldots, n\}$, 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

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

(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

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

Hellinger distance is a metric, and is related to to ℓ_1 distance in the following manner.

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

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

The square of the Hellinger distance is jointly convex.

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*

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

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 [11].

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

The chain rule for mutual information states:

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

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

The Average encoding theorem [20, 14] 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) *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 $\frac{\ln 2}{2}$.

We refer the reader to the text [21] for an introduction to the model of two-party communication protocols. We use the following Cut-and-Paste property of private-coin communication protocols (see, e.g., Ref. [5, Lemma 6.3]).

Proposition 2.5 (Cut-and-Paste) *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),*

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

2.2 The classical information cost lower bound

The 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". The notion of information on which we focus is known as "internal information" in the literature (see, e.g., Ref. [6]).

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 X, Y to the two players, Alice and Bob, respectively, are sampled from the joint distribution λ. The *information cost* of the protocol for Alice with respect to the distribution λ is defined as $\mathrm{IC}^{\mathsf{A}}_{\lambda}(\Pi) \stackrel{\text{def}}{=} \mathrm{I}(X : M \,|\, YR)$. The information cost of the protocol for Bob is defined symmetrically as $\mathrm{IC}^{\mathsf{B}}_{\lambda}(\Pi) \stackrel{\text{def}}{=} \mathrm{I}(Y : M \,|\, XR)$.

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 randmness 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 $\mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi) = \mathrm{I}(K : M^0 \,|\, XR)$. 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. Note also that under the distribution μ_0, we write Bob's input as the prefix $X[1,K]$.

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 they give her 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 $L > K$ (while maintaining the condition that $X_L = B$), 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.

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 $d \stackrel{\text{def}}{=} \mathrm{IC}^{\mathsf{A}}_{\mu_0}(\Pi)/n$, and let $c \stackrel{\text{def}}{=} \mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi)$.

To simplify the presentation, we suppress the public randomness R used in the protocol, i.e., assume that Alice and Bob only use private coins. This does not affect the generality of our proof; all the arguments below hold *mutatis mutandis* when the random variables are replaced by those conditioned on a specific value r for the public random coins R, and the parameters (d, c, ε) are replaced by the corresponding quantities $(d_r, c_r, \varepsilon_r)$. Averaging the final inequality over R and applying the Jensen Inequality gives us the claimed bound, as the inequality is of the same form as in the statement of the theorem.

Let M be the entire message transcript of the protocol. 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. This only marginally increases the information revealed by the Alice. 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. Therefore,

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

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 $\mathrm{IC}^{\mathsf{B}}_{\mu_0}(\Pi) \leq c$.

We show below that the random variables $M^0 X[1,K]$ with Bob are close in distribution to $M^1 X[1,K-1]\, \bar{X}_K$, where M^1 denotes the transcript M conditioned on the function value being 1, i.e., when $B = \bar{X}_K$.

Lemma 2.7 $\left\| M^0 X[1,K] - M^1 X[1,K-1]\, \bar{X}_K \right\| \quad \leq \quad 1 + 8\sqrt{\kappa\, c} + 4\sqrt{2\kappa\, d_1}$, *where* $\kappa = \frac{\ln 2}{2}$.

Since the protocol Π identifies the two distributions, $M^0 X[1,K]$ and $M^1 X[1,K-1]\, \bar{X}_K$, with average error ε, we have $\left\| M^0 X[1,K] - M^1 X[1,K-1]\, \bar{X}_K \right\| \geq 2(1-2\varepsilon)$. The theorem follows. ■

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: When we wish to explicitly write the transcript M as a function of the inputs to Alice and Bob, say x and $x[1,k-1], b$ respectively, we write it as $M(x; x[1,k-1], b)$. 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. Note that $M^1 = M(X; X[1,K-1], \bar{X}_K)$ has the same distribution as $M(X^{(K)}; X[1,K])$, since X and $X^{(K)}$ are identically distributed. Thus, our goal is to bound

$$\left\| M(X; X[1,K])\, X[1,K] - M(X^{(K)}; X[1,K])\, X[1,K] \right\| \ .$$

For reasons that become apparent as we develop our proof, we bound the above quantity when K is larger than $n/2$. Let L be uniformly and independently distributed in $[n] - [n/2]$. Then

$$\begin{aligned} &\left\| M(X; X[1,K])\, X[1,K] - M(X^{(K)}; X[1,K])\, X[1,K] \right\| \\ &\quad \leq \quad 1 + \frac{1}{2} \left\| M(X; X[1,L])\, X[1,L] - M(X^{(L)}; X[1,L])\, X[1,L] \right\| \ . \end{aligned} \tag{2.1}$$

So it suffices to bound the RHS 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^0 is (on average) nearly the same for different inputs to Bob.

Let J be uniformly and independently distributed in $[n/2]$, and let L be as defined above. We compare the transcript when Bob's input index is J to when it is L.

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

We defer the proof to later in this section.

When changing Alice's input, we would like to ensure that the prefix held by Bob does not change. It is for this reason that we restrict our attention to Bob's inputs with index $K \in [n/2]$, 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^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 *We have*

$$\begin{aligned}\mathbb{E}_{(x[1,l],j,l)\leftarrow(X[1,L],J,L)}\ &\mathfrak{h}(M(x[1,l]\,X[l+1,n]\,;\,x[1,j])\ ,\ M(x[1,l-1]\,\bar{x}_l\,X[l+1,n]\,;\,x[1,j]))^2\\ &\leq\quad 16\kappa\, d_1\ .\end{aligned}$$

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

$$\begin{aligned}\mathbb{E}_{(x[1,l],j,l)\leftarrow(X[1,L],J,L)}\ &\mathfrak{h}(M(x[1,l]\,X[l+1,n]\,;\,x[1,j])\ ,\ M(x[1,l]\,X[l+1,n]\,;\,x[1,l]))\\ &\leq\quad \sqrt{8\kappa\, c}\ . \end{aligned}\tag{2.2}$$

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

$$\begin{aligned}\mathbb{E}_{(x[1,l],j,l)\leftarrow(X[1,L],J,L)}\ &\mathfrak{h}(M(x[1,l]\,X[l+1,n]\,;\,x[1,l])\ ,\ M(x[1,l-1]\,\bar{x}_l\,X[l+1,n]\,;\,x[1,j]))\\ &\leq\quad \sqrt{8\kappa\, c}+\sqrt{16\kappa\, d_1}\ .\end{aligned}$$

Using the Cut-and-Paste property of 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.

$$\begin{aligned}&\mathbb{E}_{(x[1,l],j,l)\leftarrow(X[1,L],J,L)}\ \mathfrak{h}(M(x[1,l]\,X[l+1,n]\,;\,x[1,j])\ ,\ M(x[1,l-1]\,\bar{x}_l\,X[l+1,n]\,;\,x[1,l]))\\ &=\quad \mathbb{E}_{(x[1,l],j,l)\leftarrow(X[1,L],J,L)}\ \mathfrak{h}(M(x[1,l]\,X[l+1,n]\,;\,x[1,l])\ ,\ M(x[1,l-1]\,\bar{x}_l\,X[l+1,n]\,;\,x[1,j]))\\ &\leq\quad \sqrt{8\kappa\, c}+\sqrt{16\kappa\, d_1}\ .\end{aligned}\tag{2.3}$$

Combining Eq. (2.2) and Eq. (2.3), and using the Triangle Inequality we get

$$\begin{aligned}&\mathbb{E}_{(x[1,l],l)\leftarrow(X[1,L],L)}\ \mathfrak{h}(M(x[1,l]X[l+1,n]\,;\,x[1,l])\ ,\ M(x[1,l-1]\,\bar{x}_l\,X[l+1,n]\,;\,x[1,l]))\\ &\leq\quad 4\sqrt{2\kappa\, c}+4\sqrt{\kappa\, d_1}\ .\end{aligned}$$

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

$$\begin{aligned}
&\Big\| M(X\,;\,X[1,L])\,X[1,L] - M(X^{(L)}\,;\,X[1,L])\,X[1,L] \Big\| \\
&\leq \quad \mathbb{E}_{(x[1,l],j,l)\leftarrow(X[1,L],J,L)}\, \|M(x[1,l]\,X[l+1,n]\,;\,x[1,l]) - M(x[1,l-1]\,\bar{x}_l\,X[l+1,n]\,;\,x[1,l])\| \\
&\leq \quad 16\sqrt{\kappa\, c} + 8\sqrt{2\kappa\, d_1}\ .
\end{aligned}$$

The lemma follows by combining this with Eq. (2.1). ■

We return to the deferred proofs.

Proof of Lemma 2.8: Consider the random variable $\tilde{M}$ jointly distributed with X, K which is implicitly defined by the equation $KX\tilde{M} = K \otimes (XM^0)$, where the latter is the product of the two distributions K, and the marginal XM^0. Then,

Lemma 2.10 $\mathbb{E}_{(x,k)\leftarrow(X,K)}\, \mathfrak{h}\Big(M(x\,;\,x[1,k])\ ,\ \tilde{M}(x)\Big)^2 \quad \leq \quad \kappa\, c$, *where* $\kappa = \frac{\ln 2}{2}$.

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

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

which implies the lemma:

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

■

An immediate consequence of the above lemma is that

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

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

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

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

Proof of Lemma 2.9: We have

$$\mathrm{I}(X : M(X\,;\,X[1,J])\,|\,X[1,J]) \quad \leq \quad 2\,\mathrm{I}(X : M^0 | X[1,K]) \quad \leq \quad 2d_1 n\ . \tag{2.4}$$

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

$$\begin{aligned}
&\mathrm{I}(X[j+1,n] : M(x[1,j]\, X[j+1,n]\,;\, x[1,j])) \\
&\quad = \sum_{l=j+1}^{n} \mathrm{I}(X_l : M(x[1,j]\, X[j+1,n]\,;\, x[1,j]) \,|\, X[j+1,l-1]) \\
&\quad \geq \sum_{l=n/2+1}^{n} \mathrm{I}(X_l : M(x[1,j]\, X[j+1,n]\,;\, x[1,j]) \,|\, X[j+1,l-1]) \ . \qquad (2.5)
\end{aligned}$$

(2.6)

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

$$\begin{aligned}
&\mathfrak{h}(M(x[1,l-1]\, x_l\, X[l+1,n]\,;\, x[1,j]) \ ,\ M(x[1,l-1]\, \bar{x}_l\, X[l+1,n]\,;\, x[1,j]))^2 \\
&\leq \quad 4\kappa\, \mathrm{I}(X_l : M(x[1,l-1]\, X_l\, X[l+1,n]\,;\, x[1,j])) \ . \qquad (2.7)
\end{aligned}$$

Combining Eqs. (2.4), (2.5), and (2.7), we get

$$\begin{aligned}
&\mathbb{E}_{(x[1,l],j,l)\leftarrow(X[1,L],J,L)}\, \mathfrak{h}(M(x[1,l]\, X[l+1,n]\,;\, x[1,j]) \ ,\ M(x[1,l-1]\, \bar{x}_l\, X[l+1,n]\,;\, x[1,j]))^2 \\
&\leq \quad 4\kappa\, \mathbb{E}_{(x[1,l-1],j,l)\leftarrow(X[1,L-1],J,L)}\, \mathrm{I}(X_l : M(x[1,l-1]\, X_l, X[l+1,n]\,;\, x[1,j])) \\
&= \quad 4\kappa\, \mathbb{E}_{(x[1,j],j,l)\leftarrow(X[1,J],J,L)}\, \mathrm{I}(X_l : M(x[1,j]\, X[j+1,n]\,;\, x[1,j]) \,|\, X[j+1,l-1]) \\
&\leq \quad \frac{8\kappa}{n}\, \mathrm{I}(X : M(X\,;\, X[1,J]) \,|\, X[1,J]) \quad \leq \quad 16\kappa\, d_1 \ ,
\end{aligned}$$

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 given as an *input stream* $x_1, x_2, \ldots, x_n$, which arrives sequentially, i.e., letter by letter in this order. We refer the reader to the text [25] for a more thorough introduction to streaming algorithms.

Definition 3.1 (Streaming algorithm) *Fix an alphabet Σ. A 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;*
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;*
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 access to the input in the reverse order, after reaching the end of the input. Then the parameter T is the total number of passes in either direction.*

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.

The relationship between streaming algorithms for DYCK(2) and protocols for f_n is captured by a reduction due to Magniez, Mathieu, and Nayak [23]. The reduction was originally described only for one-pass streaming algorithms, but extends immediately to unidirectional multi-pass algorithms. For completeness, we sketch a proof of this theorem highlighting the differences from the one-pass case.

Theorem 3.1 (Magniez, Mathieu, and Nayak) *Any randomized streaming algorithm for* DYCK(2) *with T passes in the same direction that uses space s for instances of length $4n^2$, and has worst-case two-sided error δ yields 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: We sketch a proof of the theorem, highlighting the sole modification we need, namely in the definition of information cost. We refer the reader to Ref. [23] for the details.

We rely on the same set of hard instances of DYCK(2), which correspond to strings of length between $2n^2$ and $4n^2$. These are padded with well-formed expressions so that the length of all instances is exactly $4n^2$. Each hard instance corresponds to an instance of a $2n$-player communication protocol for ASCENSION(n), which is the logical OR of n independent instances of the two-player AUGMENTED INDEX function f_n. The players are denoted by $\mathsf{A}_i, \mathsf{B}_i$, $i \in [n]$. A T-pass unidirectional streaming algorithm for DYCK(2) that uses space s results in a communication protocol Π for ASCENSION(n) with T sequential iterations of messages in the order

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

Each message in this protocol is of length at most s, and the protocol makes the same worst-case error δ as the streaming algorithm.

Let $M_{\mathsf{B}_n,j}$, $j \in [T]$, denote the messages sent by B_n to A_n in the T iterations. The protocol Π for ASCENSION(n) gives rise to a protocol for a single instance of f_n through a direct sum property of its "internal information cost". Let μ_0 be the uniform distribution over the subset of $\big(\{0,1\}^n \times [n] \times \{0,1\}\big)$ on which the function f_n is 0. Let $(\boldsymbol{X}, \boldsymbol{k}, \boldsymbol{c}) = \big(X^i, k^i, c^i\big)_{i=1}^n$ be n instances of f_n, distributed according to μ_0^n. Let R denote the public random bits in the protocol Π arising from the randomness used by the streaming algorithm. The (internal) information cost of Π is defined as:

$$\mathrm{IC}_{\mu_0^n}(\Pi) \quad = \quad \mathrm{I}(\boldsymbol{k}, \boldsymbol{c} : M_{\mathsf{B}_n,1} \cdots M_{\mathsf{B}_n,T} \,|\, \boldsymbol{X} R) \ .$$

This is the natural and straightforward extension of the measure used in the one-pass case, which concentrates on $M_{\mathsf{B}_n,1}$, the single message sent by B_n. Note that $\mathrm{IC}_{\mu_0^n}(\Pi) \leq Ts$, as each message $M_{\mathsf{B}_n,j}$ is of length at most s.

The protocol Π may be adapted to n different protocols Π'_i, $i \in [n]$, for f_n, by precisely the same method of embedding an instance (X, K, B) of f_n into one of ASCENSION(n), as described in Ref. [23, Section 4.3]. The $2n$ players in Π are simulated by two players, Alice and Bob, as before: Alice simulates $\mathsf{A}_1, \mathsf{B}_1, \mathsf{A}_2, \mathsf{B}_2, \ldots, \mathsf{A}_i$, sends a message to Bob, who simulates $\mathsf{B}_i, \mathsf{A}_{i+1}, \mathsf{B}_{i+1}, \ldots, \mathsf{A}_n, \mathsf{B}_n$,

sends a message to Alice, who simulates $\mathsf{A}_n, \mathsf{A}_{n-1}, \ldots, \mathsf{A}_1$, and they repeat this in the same order a total of T times. There are $2T$ messages in this protocol starting with Alice, she uses only public randomness, whereas Bob may use private randomness, and the protocol makes the same distributional error (at most δ) on the uniform distribution over its inputs as Π does. The (internal) information cost of Π'_i is measured as

$$\mathrm{IC}_{\mu_0}(\Pi'_i) \quad = \quad \mathrm{I}(k^i, c^i : M_{\mathsf{B}_n,1} \cdots M_{\mathsf{B}_n,T} \,|\, X^i R^i) \ ,$$

where R^i is the public randomness in Π'_i. This is the mutual information of all the messages sent by Bob with his input, given Alice's input, under the uniform distribution over the 0s of the function f_n.

The superadditivity of mutual information gives us the direct sum result

$$\mathrm{IC}_{\mu_0^n}(\Pi) \quad = \quad \sum_{i=1}^{n} \mathrm{IC}_{\mu_0}(\Pi'_i),$$

as in Ref. [23, Lemma 3]. Therefore at least one protocol for f_n from (Π'_i), call it Π', has internal information cost at most Ts/n. Note that we may replace Bob's messages by the entire message transcript in Π' in this information cost without changing its value, as Alice's messages are independent of K, given X, Bob's messages, and the public randomness. Moreover, the total length of the messages sent by Alice is at most sT, so the mutual information of X with the entire message transcript in Π', even given Bob's input and the public randomness, is at most sT. ■

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

$$\frac{\sqrt{N}}{T} \times \frac{1}{6+4\sqrt{2}} \left[\frac{1-4\varepsilon}{4\sqrt{\ln 2}} - \frac{2\sqrt{\mathrm{H}(2\varepsilon)}}{\sqrt[4]{N}} \right]^2$$

on instances of length N.

4 Quantum information cost of Augmented Index

We now turn to quantum communication, and present the necessary background in Section 4.1. In Section 4.2, 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 Indexinvolves a number of additional additional subtleties, which are also described along the way.

4.1 Quantum information theory and communication

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. 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 Bures distance $\mathfrak{h}(A\ ,\ B)$ between the states is defined as

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

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 to ℓ_1 distance in the following manner.

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 square of the Bures distance is convex in the following sense. Suppose two quantum states P, Q are block diagonal in the same basis $|x\rangle$ for the space $\mathbb{C}^{\mathcal{S}}$, and the blocks corresponding to x in P, Q have the same trace p_x.

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*

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

The Local Transition Theorem due to Uhlmann [27] 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 $|\psi_1\rangle$ 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 [27].

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 $XPQ = \sum_{x\in\mathcal{S}} p_x|x\rangle\langle x| \otimes P_xQ_x$, where P_xQ_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 $XYQ = \sum_{x\in\mathcal{S},y\in\mathcal{S}'} p_xq_y|xy\rangle\langle xy| \otimes Q_{xy}$ be a joint quantum state. Then*

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

The Average Encoding Theorem [20, 14] 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 $\frac{\ln 2}{2}$.

We briefly describe the model of two-party quantum communication that we study. Following the model introduced by Yao [29], two "players", Alice and Bob, hold some number of qubits, which initially factor into a tensor product $\mathcal{A} \otimes \mathcal{H}_{\mathsf{A},0} \otimes \mathcal{H}_{\mathsf{B},0} \otimes \mathcal{B}$ of Hilbert spaces. The qubits corresponding to $\mathcal{A} \otimes \mathcal{H}_{\mathsf{A},0}$ are in Alice's possession, and those in $\mathcal{H}_{\mathsf{B},0} \otimes \mathcal{B}$ are in Bob's possession. We restrict ourselves to protocols with classical inputs and outputs. When the game starts, Alice holds a classical input represented by a bit string x and similarly Bob holds y. In other words, 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 spaces $\mathcal{H}_{\mathsf{A},0} \otimes \mathcal{H}_{\mathsf{B},0}$ are intended to be the workspace of the two parties, and are initialized to a possibly entangled state $|\Phi\rangle$ that is independent of the inputs Alice and Bob have. The initial joint state is thus $|x\rangle \otimes |\Phi\rangle \otimes |y\rangle$.

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 in round i, with $i \geq 1$. Suppose 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 unitary depends on her input x and the round. We will have occasion to consider runs of the protocol on superpositions of inputs. In this case, we think of Alice as applying the unitary $\sum_x |x\rangle\langle x| \otimes V_{i,x}$ to the qubits in the space $\mathcal{A} \otimes \mathcal{H}_{\mathsf{A},i-1}$. Then, Alice sends some of her qubits to Bob. Formally, the space $\mathcal{H}_{\mathsf{A},i-1}$ factors as $\mathcal{H}_{\mathsf{A},i} \otimes \mathcal{M}_i$, where $\mathcal{M}_i$ denotes the message space, and $\mathcal{H}_{\mathsf{B},i} = \mathcal{M}_i \otimes \mathcal{H}_{\mathsf{B},i-1}$. As a result, Bob may now apply a unitary operation to the 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, measures the qubits in his possession (those in $\mathcal{H}_{\mathsf{B},t}$) according to a general measurement that may depend on his input y. The measurement outcome is considered to be the output of the protocol.

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.

4.2 The quantum information cost trade-off

In this section, we derive an analogue of the information trade-off result established in Section 2.2 for quantum communication protocols for Augmented Index.

We first define the notion of *quantum* information cost for the Augmented Index function f_n. As in Section 2.2, 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.

A significant difference between the classical and quantum information costs arises because the no-cloning principle [27] 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. A second point of departure from the classical case is that we consider the information contained about a *superposition* of inputs corresponding to the distribution of interest. This information is in general more than the information contained about a distribution over inputs, and the resulting notion seems to be necessary for the proof of the information cost trade-off we present. The final, technical point of difference comes from the manner in which the input is distributed among the two parties. Since Alice and Bob share $X[1, K]$, when the input registers are initialized with superpositions corresponding to μ_0, the two parties already begin with some information about the each other's input. Unlike in the classical case, this enables Alice to get information about the index K. The effect of sharing the prefix $X[1, K]$ is identical to that of measuring the first K qubits of Alice's superposition in the computational basis. This results in states of varying amount of von Neumann entropy for different indices, which leaks information about the index K. To quantify the information leaked by the protocol, we therefore imagine that there is a single quantum register that carries the superposition corresponding to X, and that Bob has read-only access to the relevant portion of this register. The information cost is then measured with respect to this register.

As explained above, we adopt the following convention with respect to the inputs for AUGMENTED INDEX in the rest of this section. We imagine that Alice is given the input x, and 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 the local unitary operations used by Bob during the protocol are controlled by the register holding this prefix.

We suppose that there are a total of t messages, beginning with Alice and alternating with Bob. This is solely to eliminate awkwardness in defining and referring to quantum information cost as we do below, and may be removed without affecting the results. 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 $P_i Q_i$ denote the joint state of Alice and Bob's workspace in a protocol Π for AUGMENTED INDEX when we start with a uniform superposition $\hat{X}$ over string $x \in \{0,1\}^n$ and the random inputs K, B with Bob (this corresponds to distribution μ), and let $P_i^0 Q_i^0$ denote the analogous joint state corresponding to μ_0, immediately after the ith message is sent. The quantum information cost of Π for Alice and Bob with respect to μ_0 is then defined as

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

Note that there is an asymmetry in the manner we quantify quantum information cost. In Alice's cost, we measure the information about a uniformly random string X in Bob's quantum state, given the prefix to which he has access. In Bob's cost, we measure the information about a random index K in the joint state of strings x in superposition and Alice's workspace qubits. Although we could also consider superpositions over x in Alice's cost and over k in Bob's cost, we chose the above notions as they give us the strongest result. The information quantities with superpositions are always bounded from below by the ones with random variables, due to the monotonicity of mutual information under quantum operations.

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. In the final piece of the argument above, 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 [14], 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.

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*

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

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 $d \stackrel{\mathrm{def}}{=} \mathrm{QIC}^{\mathsf{A}}_{\mu_0}(\Pi)/n$, and let $c \stackrel{\mathrm{def}}{=} \mathrm{QIC}^{\mathsf{B}}_{\mu_0}(\Pi)$.

Let $\hat{X} P_i Q_i K B$ 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 $d_i = \frac{1}{n}\,\mathrm{I}(X : Q_i^0 \,|\, X[1,K])$ for odd $i \in [t]$, and $c_i = \mathrm{I}(K : \hat{X} P_i^0)$ 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 $\left\| XP_t^0 - XP_t^1 \right\|_{\mathrm{tr}} \quad \leq \quad 1 + 4\sqrt{\kappa\, t}\left[2\sqrt{d} + \sqrt{2c}\right], \quad$ *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, we have

$$\left\| XP_t^0 - XP_t^1 \right\| \quad \geq \quad 2(1-2\varepsilon) \ .$$

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 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

$$\begin{aligned} & \left\| XP_t(X; X[1,K]) - X^{(K)} P_t(X^{(K)}; X[1,K]) \right\|_{\mathrm{tr}} \\ \leq \;& 1 + \frac{1}{2} \left\| XP_t(X; X[1,J]) - X^{(L)} P_t(X^{(L)}; X[1,L]) \right\| \\ = \;& 1 + \frac{1}{2} \left\| X^{(L)} P_t(X^{(L)}; X[1,J]) - X^{(L)} P_t(X^{(L)}; X[1,L]) \right\| . \end{aligned} \tag{4.1}$$

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

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

This expression 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 $\hat{X}[l+1,n]\, P_i Q_i$, 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	$j, z[1,j-1], z_j$	$\lvert\phi_i(z,j)\rangle$
01	z	$l, z[1,l-1], z_l$	$\lvert\phi_i(z,l)\rangle$
10	$z^{(l)}$	$j, z[1,j-1], z_j$	$\lvert\phi_i(z^{(l)},j)\rangle$
11	$z^{(l)}$	$l, z[1,l-1], z_l$	$\lvert\phi_i(z^{(l)},l)\rangle$

The "Run" column indicates whether Alice's lth bit has been switched, and whether we have switched j to l. Note that in the first three runs of the protocol, we expect the output to be 0, and in the last run, we expect it to be 1.

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

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

for every odd $i \in [t]$. Define

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

for every even $i \in [t]$. 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 carefully chosen 0-inputs of Augmented Index to corresponding 1-inputs 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.

By the Triangle Inequality, the monotonicity of the trace distance under quantum operations, the relationship between trace and Bures distance (Proposition 4.1), Lemmata 4.10, 4.8 and 4.9,

$$\begin{aligned}
&\Big\|X^{(L)}P_t(X^{(L)};X[1,J]) - X^{(L)}P_t(X^{(L)};X[1,L])\Big\|_{\mathrm{tr}} \\
&\quad\leq\ \mathbb{E}_{(j,l,z)\leftarrow(J,L,X[1,L])}\ \Big\|X[l+1,n]\,P_t(z^{(l)}X[l+1,n];z[1,j]) - X[l+1,n]\,P_t(z^{(l)}X[l+1,n];z[1,l])\Big\|_{\mathrm{tr}} \\
&\quad\leq\ \mathbb{E}_{(j,l,z)\leftarrow(J,L,X[1,L])}\ \Big\|\hat{X}[l+1,n]\,P_t(z^{(l)}\hat{X}[l+1,n];z[1,j]) - \hat{X}[l+1,n]\,P_t(z^{(l)}\hat{X}[l+1,n];z[1,l])\Big\|_{\mathrm{tr}} \\
&\quad\leq\ 2\sqrt{2}\ \mathbb{E}_{(j,l,z)\leftarrow(J,L,X[1,L])}\ \mathfrak{h}\Big(\hat{X}[l+1,n]\,P_t(z^{(l)}\hat{X}[l+1,n];z[1,j])\ ,\ \hat{X}[l+1,n]\,P_t(z^{(l)}\hat{X}[l+1,n];z[1,l])\Big) \\
&\quad\leq\ 2\sqrt{2}\ \mathbb{E}_{(j,l,z)\leftarrow(J,L,X[1,L])}\ \mathfrak{h}\Big((\mathbb{I}_{\mathsf{A}} \otimes U_t)|\phi_t(z^{(l)},j)\rangle\ ,\ |\phi_t(z^{(l)},l)\rangle\Big) \\
&\quad\leq\ 4\sqrt{2}\ \mathbb{E}_{(j,l,z)\leftarrow(J,L,X[1,L])}\ \sum_{i=1}^{t} h_i(j,l,z) \\
&\quad\leq\ 4\sqrt{2}\left[\sum_{\text{odd } i\in[t]} 4\sqrt{\kappa\, d_i} + \sum_{\text{even } i\in[t]} 2\sqrt{2\kappa\, c_i}\right] \quad\leq\quad 8\sqrt{\kappa\, t}\left[2\sqrt{d} + \sqrt{2c}\right]\ .
\end{aligned}$$

This concludes the proof of Lemma 4.7. ∎

We turn to the deferred proofs.

Proof of Lemma 4.8: 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 $j' \in [n/2]$, $l' \in [n]-[n/2]$,

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

Since Bures distance is monotonic under measurements, measuring the first l' qubits of $\hat{X}$ yields

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

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. ∎

Proof of Lemma 4.9: 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

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

We have

$$\mathrm{I}(X : Q_i^0(X\,;\,X[1,J])\,|\,X[1,J]) \ \leq\ 2\,\mathrm{I}(X : Q_i^0\,|\,X[1,K]) \quad \leq \quad 2d_i n \ . \tag{4.2}$$

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

$$\begin{aligned}
&\mathrm{I}(X[j'+1,n] : Q_i(z''X[j'+1,n]\,;\,z'')) \\
&\quad = \sum_{l'=j'+1}^{n} \mathrm{I}(X_{l'} : Q_i(z''X[j'+1,n]\,;\,z'') \,|\, X[j'+1,l'-1]) \\
&\quad \geq \sum_{l'=n/2+1}^{n} \mathrm{I}(X_{l'} : Q_i(z''X[j'+1,n]\,;\,z'') \,|\, X[j'+1,l'-1]) \ . \qquad (4.3)
\end{aligned}$$

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

$$\begin{aligned}
&\mathfrak{h}\Big(Q_i(z'X[l'+1,n]\,;\,z'[1,j'])\ ,\ Q_i(z'^{(l')}X[l'+1,n]\,;\,z'[1,j'])\Big)^2 \\
&\leq \quad 4\kappa\ \mathrm{I}(X_{l'} : Q_i(z'[1,l'-1]\,X_{l'}\,X[l'+1,n]\,;\,z'[1,j'])) \ . \qquad (4.4)
\end{aligned}$$

Combining Eqs. (4.2), (4.3), and (4.4), 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. ∎

Proof of Lemma 4.10: 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

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

so along with Lemma 4.9 we get

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

We prove that the lemma holds for r, assuming that it holds for $r-1 \in [t]$. 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.

By Lemma 4.8, we have

$$\mathfrak{h}(\,(\mathbb{I}_{\mathsf{A}} \otimes U_r)\,|\phi_r(z,j)\rangle\ ,\ |\phi_r(z,l)\rangle) \quad = \quad h_r(j,l,z) \ , \qquad (4.5)$$

and by Lemma 4.9 we have

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

By the induction hypothesis, we also have

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

Now

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

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 unitary operators, we get

$$\mathfrak{h}\Big((U_{r-1}\otimes\mathbb{I}_{\mathsf{B}})\,|\phi_r\rangle(z,j)\ ,\ |\phi_r(z^{(l)},j)\rangle\Big)\ \ =\ \ h_{r-1}(j,l,z)\ , \tag{4.6}$$

and

$$\mathfrak{h}\Big((U_{r-1}\otimes\mathbb{I}_{\mathsf{B}})\,|\phi_r(z,l)\rangle\ ,\ |\phi_r(z^{(l)},l)\rangle\Big)\ \ \leq\ \ h_{r-1}(j,l,z)+2\sum_{i=1}^{r-2}h_i(j,l,z)\ . \tag{4.7}$$

Using Eqs. (4.5), (4.6), and (4.7), and observing that U_{r-1} and U_r act on disjoint sets of qubits, we get

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

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

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. [7] for sending us their initial manuscript when we first publicized an earlier version of the classical results in this article. The 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.

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] Mark Braverman and Anup Rao. Efficient communication using partial information. Technical Report TR10-083, Electronic Colloquium on Computational Complexity, `http://eccc.hpi-web.de/`, May 13 2010.

[7] Amit Chakrabarti, Ranganath Kondapally Graham Cormode, and Andrew McGregor. Information cost tradeoffs for Augmented Index and streaming language recognition. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science*, pages 387–396, Washington, DC, USA, 2010. IEEE Computer Society.

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

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

[20] 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.

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

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

[23] 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.

[24] 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.

[25] 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.

[26] 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.

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

[28] 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.

[29] 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.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de