Electronic Colloquium on Computational Complexity, Revision 1 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 4, 2010

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 the peculiar power of bi-directional streams over unidirectional ones observed in the algorithms they present.

The lower bound is obtained by analysing 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.

*Centre for Quantum Technologies and Department of Computer Science, National University of Singapore. rahul@comp.nus.edu.sg

†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 the Center for Quantum Technologies, National University of Singapore. Research supported in part by NSERC Canada. 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

The language DYCK(2) consists of all well-parenthesized expressions over two types of parenthesis, denoted below by $a, \overline{a}$ and $b, \overline{b}$.

Definition 1.1 DYCK(2) *is the language over alphabet* $\Sigma = \{a, \overline{a}, b, \overline{b}\}$ *defined recursively by:*

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

This deceptively simple language is complete for the class of context-free languages [6], and is implicit in a myriad of information processing tasks. It has been studied extensively, most recently in setting of streaming algorithms [14]. Streaming algorithms are designed with the idea of processing massive data, which cannot fit entirely in computer memory. Consequently, random access to the input is extremely expensive, and furthermore, the algorithms are required 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. (See the text [15] for an introduction to this topic.)

Magniez, Mathieu, and Nayak [14] present two 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. They show that the space requirement 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 startling property of the second algorithm is that it makes the second pass in *reverse* order, and this seems essential for its performance. This phenomenon is partially explained by the authors, by way of a space lower bound for one-pass algorithms. They prove that any one-pass algorithm that makes error at most $1/n \log n$ uses space $\Omega(\sqrt{n \log n})$, and conjecture that a similar bound hold for multi-pass streaming algorithms if all passes are made in the same direction.

Logarithmic space is sufficient to recognize DYCK(2), if we are allowed random access to the input: we may run through all possible heights, and check parentheses at the same height. This scheme translates to streaming algorithms with a linear number of passes, but does not rule out the possibility of algorithms with fewer (but more than one) passes, that use sub-polynomial space. We show an $\Omega(\sqrt{n}/T)$ lower bound for the space required by any unidirectional randomized T-pass streaming algorithm that recognizes DYCK(2) with a constant probability of error. This proves the conjecture from [14] and establishes the peculiar power of bi-directional streams.

A relatively straightforward generalization of the one-pass algorithm in [14] gives us a unidirectional randomized T-pass streaming algorithm that uses space $O\left(\sqrt{(n \log n)/T}\right)$ and has polynomially small probability of error. The lower bound we derive thus comes within a factor $\sqrt{\log n}/T^{1/2}$ of optimal. The bound for one pass algorithms is a factor of $\sqrt{\log n}$ better than the one in [14], for constant error probability, but falls short of optimal (by the same factor) for polynomially small error.

We derive the above lower bound by following the same high level route as taken in [14]. They map a streaming algorithm with space s for DYCK(2) to a multi-party communication protocol in which the messages are each of the same length s, and then bound s from below through a communication complexity bound. The communication bound is derived using the information cost approach (see, for example, [5, 17, 2, 11, 9]), which reduces the task to bounding from below the information cost

of a variant of the INDEX problem, in which 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 first studied in the one-way communication model as "serial encoding" [1, 16], and is called the "Mountain problem" in [14].

It is in the analysis of the information cost of f_n that we depart from the earlier route. First, we formulate a bound for protocols resulting from streaming algorithms entirely in information-theoretic terms. Second, we demonstrate 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 this variant of INDEX, in spite of the apparent differences from earlier 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 $f_n^{-1}(0)$).

A notion of information cost for INDEX has been studied previously by Jain, Radhakrishnan, and Sen [10] in the context of privacy in communication. This notion, defined in terms of the hard (product) distribution for the problem, however seems not to be directly relevant to our situation, where we deal with an easy and non-product distribution.

In independent work, concurrent with an earlier version of this article [8], Chakrabarti, Cormode, Kondapally, and McGregor [4] derived a similar information cost trade-off for f_n, for protocols that make error $o(1/\log n)$ on the uniform distribution μ. While the basic tools from information theory that they ultimately employ are similar to ours, they take a different, rather technical, route to these tools. Their motivation is identical to ours—to study the space required by unidirectional multi-pass streaming algorithms for DYCK(2). Their result for f_n leads to an $\Omega(\sqrt{n \log\log n}/T)$ space lower bound for unidirectional T-pass randomized streaming algorithms for DYCK(2) that make constant error, a factor of $\Omega(\sqrt{\log\log n}\,)$ smaller than the one we conclude. There is a subtle but significant difference between our results on the information cost trade-off for f_n themselves. Parallel repetition, the only known method for amplifying the success probability in this context, potentially blows up the information revealed by the two parties. Trade-offs for protocols that make smaller error therefore do not imply any trade-offs for constant error protocols.

In the first version of this article [8], we derived a trade-off for constant error protocols which depends on T, the number of rounds of message exchange the protocols have. We showed that either Alice reveals $\Omega(n/T)$ information about her input x, or Bob reveals $\Omega(1/T)$ information about k (when these inputs are drawn from the distribution μ_0). This leads to an $\Omega(\sqrt{n}/T^3)$ space lower bound for unidirectional T-pass randomized streaming algorithms for DYCK(2) that make constant error. This is smaller by a factor of T^2 than the bound we present here. It outperforms the one in [4] by a factor of $\Omega(\sqrt{\log\log n}\,)$ when T is constant, but this advantage shrinks in the obvious manner as T grows. It is overtaken by the latter bound when $T \in \Omega(\sqrt[4]{\log\log n}\,)$. Notwithstanding the dependence on the number of rounds, our remarks about trade-offs for constant error protocols vis-a-vis those for protocols making smaller error (as in [4]) still apply.

The intuition behind the proof in the current version of our article is the same as that in the original, and may be understood in more detail in Section 2.3. The original proof relied on a round reduction type of argument giving rise to the dependence on the number of rounds mentioned above. Our innovation here is to realize the same intuition in one-shot, by analysing the protocol transcript as a whole, instead of analysing it round-by-round.

2 Lower bounds for unidirectional streams

In this section we present the results of this article. We derive a bound on the space required by streaming algorithms for DYCK(2). The lower bound is derived by invoking methods from communication complexity. We use the reduction due to Magniez, Mathieu, and Nayak [14] between a streaming algorithm for DYCK(2) and two-party communication protocols for the variant f_n of INDEX. While their connection is described for one-pass algorithms, it holds *mutatis mutandis* for (unidirectional) multi-pass streaming algorithms. We state this connection in Section 2.2, and then develop our lower bound in Section 2.3. We summarize the notational conventions we follow and the background from information theory that we assume in Section 2.1.

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 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 $i, j \in [n]$, 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 $z \leftarrow Z$. We denote the ℓ_1-distance between two random variables A, B over the same sample space by $\|A - B\|$, and the Hellinger distance between them as $\mathfrak{h}(A\ ,\ B)$.

We rely on a number of standard facts from information theory in this work. For a comprehensive introduction to information theory, we refer the reader to a text such as [7].

The Hellinger distance is a metric, and the following fact relates it to ℓ_1 distance.

Fact 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.

Fact 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*

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

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 says:

Fact 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 [12, 9] is a quantitative version of the fact 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.

Fact 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 need the following Cut-and-Paste property of two-party private-coins communication protocols (see e.g. [2, Lemma 6.3]). We refer the reader to the text [13] for an introduction to the model of two-party communication protocols.

Fact 2.5 (Cut-and-Paste) *Let Π be a two-party private coins communication protocol. Let $M(x, y)$ denote the random variable representing the message transcript in Π when Alice has input x and Bob has input y. Then,*

$$\mathfrak{h}(M(x_1, y_1) \ , \ M(x_2, y_2)) \quad = \quad \mathfrak{h}(M(x_1, y_2) \ , \ M(x_2, y_1)) \ .$$

2.2 The two-party communication problem

We consider randomized two-party communication protocols arising from streaming algorithms for Dyck languages. As described in Section 1, in these protocols 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 variant of the index function problem is called the "Mountain problem" in [14], the "Augmented Index problem" in [4], and was previously studied in the setting of one-way communication as "serial encoding" [1, 16].

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. Let M denote the entire message transcript under μ, and let M^0 denote the transcript under distribution μ_0. The protocols for f_n on which we focus satisfy the following properties that arise from considerations in the streaming model.

1. Alice and Bob may use private randomness in addition to public randomness R.
2. The information that the transcript carries about X under distribution μ_0, from Bob's point of view, is small:
$$\mathrm{I}(X : M^0 | \, X[1, K] \, R) \quad \leq \quad dn \ ,$$
for some $d \geq 0$. As K can be inferred from the prefix $X[1, K]$, we have suppressed it in the conditioning.

3. The information that the transcript carries about K under distribution μ_0 is small from Alice's point of view:
$$\mathrm{I}(K : M^0|XR) \quad \leq \quad c \ ,$$
for some $c \geq 0$.
4. The distributional error of the protocol under the uniform distribution μ over inputs is at most $\varepsilon < \frac{1}{2}$.

We refer to protocols as described above as (d, c, ε)-protocols for the function f_n.

A less symmetric notion of information, $\mathrm{I}(X : M^0|KR)$, is considered by Chakrabarti *et al.* [4] instead of the one we have in Item 2 above. The condition in Item 2 from which we start is a weaker constraint on protocols than a bound on $\mathrm{I}(X : M^0|KR)$, as

$$\mathrm{I}(X : M^0|\, X[1,K]\, R) \quad \leq \quad \mathrm{I}(X : M^0|\, X[1,K]\, R) + \mathrm{I}(X[1,K] : M^0|\, KR) \quad = \quad \mathrm{I}(X : M^0|\, KR) \ .$$

The notion of information which we choose to study is arguably more natural, and has occurred in previous works under the name "internal information" (see, e.g., [3]). The previous version of this article [8] assumed a bound on the total message length. However, the proof technique we used relied only on a bound on the information $\mathrm{I}(X : M^0|R)$. This kind of bound is also a weaker requirement than one on $\mathrm{I}(X : M^0|KR)$, as

$$\mathrm{I}(X : M^0|R) \quad \leq \quad \mathrm{I}(X : M^0K|R) \quad = \quad \mathrm{I}(X : M^0|KR) \ ,$$

as X and K are independent, and independent of R. However, it is not much weaker. Since

$$\mathrm{I}(X : M^0|KR) \quad \leq \quad \mathrm{I}(X : M^0|R) + \log_2 n \ ,$$

a bound of dn, with $d = \Omega\left(\frac{\log n}{n}\right)$, on $\mathrm{I}(X : M^0|R)$ implies a similar bound on $\mathrm{I}(X : M^0|KR)$. We note however, that the analysis of the index function in [4] may be adapted to work under the condition in Item 2.

The relationship between streaming algorithms and protocols for f_n is captured by the following reduction.

Theorem 2.6 (Magniez, Mathieu, and Nayak [14]) *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 implies an $(sT/n, sT/n, \delta)$ streaming protocol for f_n.*

The reduction was described in [14] only for one-pass streaming algorithms, but extends immediately to multi-pass algorithms. For completeness, we sketch a proof of this theorem in Appendix A, highlighting the differences from the one-pass case.

2.3 The communication 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 f_n. We show that at least one of the parties necessarily reveals "a lot" of information even on the "easy distribution" μ_0 if the protocol computes f_n with bounded error on a "hard distribution" μ. We state the theorem for even n. A qualitatively similar result holds for odd n, and may be derived from the proof for the even case.

Theorem 2.7 *Any (d, c, ε)-protocol for f_n with $\varepsilon \leq 1/4$ and n even satisfies*

$$4\sqrt{c} + 2\sqrt{d} \quad \geq \quad \frac{(1-4\varepsilon)}{\sqrt{2\ln 2}} - 4\sqrt{\mathrm{H}(2\varepsilon)} \ .$$

Proof: Consider a (d, c, ε)-protocol P for f_n, in which Alice's input X is uniformly distributed over $\{0,1\}^n$, and Bob's input is $(K, X[1, K-1], B)$, where K, B are uniformly distributed over $[n]$ and $\{0,1\}$, respectively. In addition, X, K, B are all independent.

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 the output of the protocol may be computed from M. This may be accomplished by including the output in the final message, while only marginally increasing the information revealed by the party sending the final message. 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, if Alice sends this output bit,

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

and $\mathrm{I}(K : M^0 O^0 |\, X) = \mathrm{I}(K : M^0 |\, X)$. If Bob sends the output bit, then

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

and $\mathrm{I}(X : M^0 O^0 |\, X[1,K]) = \mathrm{I}(X : M^0 |\, X[1,K])$. Henceforth, we assume that P is a (d_1, c_1, ε)-protocol in which the output may be computed from M, and either $d_1 = d + \mathrm{H}(2\varepsilon)/n, c_1 = c$, or $d_1 = d, c_1 = c + \mathrm{H}(2\varepsilon)$.

We show below that the message transcript M^0 is close in distribution to the message transcript M^1, which denotes the transcript M conditioned on the function value being 1, i.e., when $B = \bar{X}_K$.

Lemma 2.8 $\left\| M^0 - M^1 \right\| \quad \leq \quad 1 + 8\sqrt{\kappa\, c_1} + 2\sqrt{2\kappa\, d_1}$, *where* $\kappa = \frac{\ln 2}{2}$.

Since the protocol P identifies the two distributions, M^0 and M^1, with average error ε, we have $\left\| M^0 - M^1 \right\| \geq 2(1 - 2\varepsilon)$. The theorem follows. ∎

This immediately gives us a space lower bound for one-pass streaming algorithms for Dyck(2). Let δ_0 be the root in $[0, 1/2]$ of the function $E(z)$ defined as

$$E(z) \quad = \quad \frac{(1-4z)}{\sqrt{2\ln 2}} - 4\sqrt{\mathrm{H}(2z)} \ .$$

Corollary 2.9 *Any randomized T-pass streaming algorithm for* DYCK(2) *that has worst case two-sided error $\delta < \delta_0$ uses space at least*

$$\frac{E(\varepsilon)^2}{72} \times \frac{\sqrt{N}}{T}$$

on instances of length N.

Proof of Lemma 2.8: 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

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

Let J be uniformly and independently distributed in $[n/2]$, and let L be uniformly and independently distributed in $[n] - [n/2]$. Then

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

So it suffices to bound the RHS above.

Since it does not carry much information about K, we deduce that the transcript M^0 does not distinguish between different inputs to Bob.

Lemma 2.10 $\mathbb{E}_{(x,j,l) \leftarrow (X,J,L)} \, \mathfrak{h}(M(x\,;\, x[1,j]) \ , \ M(x\,;\, x[1,l]))^2 \quad \leq \quad 8\kappa\, c_1.$

We defer the proof to later in this section.

Since M^0 does not carry much information about X, even given a prefix, flipping a bit outside the prefix does not perturb it by much.

Lemma 2.11 *We have*

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

This is proven later in the section.

We now conclude the proof of Lemma 2.8. Since Hellinger distance squared is jointly convex (Fact 2.2), Lemma 2.10 gives us

$$\begin{aligned} \mathbb{E}_{(x[1,l],j,l) \leftarrow (X[1,L],J,L)} \, &\mathfrak{h}(M(x[1,l]\, X[l+1,n]\,;\, x[1,j]) \ , \ M(x[1,l]\, X[l+1,n]\,;\, x[1,l])) \\ &\leq \quad \sqrt{8\kappa\, c_1} \ . \end{aligned} \tag{2.2}$$

Along with the triangle inequality, and Lemma 2.11, this implies that

$$\begin{aligned} \mathbb{E}_{(x[1,l],j,l) \leftarrow (X[1,L],J,L)} \, &\mathfrak{h}(M(x[1,l]\, X[l+1,n]\,;\, x[1,l]) \ , \ M(x[1,l-1]\, \bar{x}_l\, X[l+1,n]\,;\, x[1,j])) \\ &\leq \quad \sqrt{8\kappa\, c_1} + \sqrt{4\kappa\, d_1} \ . \end{aligned}$$

Using the Cut-and-Paste property of communication protocols (Fact 2.5), we get

$$\begin{aligned}
&\mathbb{E}_{(x[1,l],j,l)\leftarrow(X[1,L],J,L)}\ \mathfrak{h}(M(x[1,l]\,X[l+1,n]\,;\,x[1,j])\ ,\ M(x[1,l-1]\,\bar{x}_l\,X[l+1,n]\,;\,x[1,l])) \\
&= \quad \mathbb{E}_{(x[1,l],j,l)\leftarrow(X[1,L],J,L)}\ \mathfrak{h}(M(x[1,l]\,X[l+1,n]\,;\,x[1,l])\ ,\ M(x[1,l-1]\,\bar{x}_l\,X[l+1,n]\,;\,x[1,j])) \\
&\leq \quad \sqrt{8\kappa\,c_1}+\sqrt{4\kappa\,d_1}\ . \qquad (2.3)
\end{aligned}$$

Combining Eq. (2.2) and Eq. (2.3), and using the triangle inequality we get

$$\begin{aligned}
&\mathbb{E}_{(x[1,l],l)\leftarrow(X[1,L],L)}\ \mathfrak{h}(M(x[1,l]X[l+1,n]\,;\,x[1,l])\ ,\ M(x[1,l-1]\,\bar{x}_l\,X[l+1,n]\,;\,x[1,l])) \\
&\leq \quad 4\sqrt{2\kappa\,c_1}+2\sqrt{\kappa\,d_1}\ .
\end{aligned}$$

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

$$\begin{aligned}
&\Big\|M(X\,;\,X[1,L])-M(X^{(L)}\,;\,X[1,L])\Big\| \\
&= \quad \mathbb{E}_{(x[1,l],j,l)\leftarrow(X[1,L],J,L)}\ \|M(x[1,l]\,X[l+1,n]\,;\,x[1,l])-M(x[1,l-1]\,\bar{x}_l\,X[l+1,n]\,;\,x[1,l])\| \\
&\leq \quad 16\sqrt{\kappa\,c_1}+4\sqrt{2\kappa\,d_1}\ .
\end{aligned}$$

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

We return to the deferred proofs.

Proof of Lemma 2.10: Define a random variable $\tilde{M}$ implicitly 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 2.12 $\mathbb{E}_{(x,k)\leftarrow(X,K)}\ \mathfrak{h}\Big(M(x\,;\,x[1,k])\ ,\ \tilde{M}(x)\Big)^2 \quad\leq\quad \kappa\,c_1$, *where* $\kappa=\frac{\ln 2}{2}$.

Proof: From the average encoding theorem, Fact 2.4, we have that for every $x\in\{0,1\}^n$,

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

which implies the lemma:

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

■

An immediate consequence of the above lemma is that

$$\begin{aligned}
\mathbb{E}_{(x,j)\leftarrow(X,J)}\ \mathfrak{h}\Big(M(x\,;\,x[1,j])\ ,\ \tilde{M}(x)\Big)^2 &\leq 2\,\kappa\,c_1\ , \qquad \text{and} \\
\mathbb{E}_{(x,l)\leftarrow(X,L)}\ \mathfrak{h}\Big(M(x\,;\,x[1,l])\ ,\ \tilde{M}(x)\Big)^2 &\leq 2\,\kappa\,c_1\ .
\end{aligned}$$

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 2.11: We have

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

Fix a sample point $x[1,j], j$. By the chain rule (Fact 2.3),

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

Moreover, by the average encoding theorem (Fact 2.4), for any given $x[1,l]$,

$$\begin{aligned}
&\mathfrak{h}(M(x[1,l-1]\, x_l\, X[l+1,n]\,;\, x[1,j])\ ,\ M(x[1,l-1]\, \bar{x}_l\, X[l+1,n]\,;\, x[1,j]))^2 \\
&\leq \quad \kappa\, \mathrm{I}(X_l : M(x[1,l-1]\, X_l\, X[l+1,n]\,;\, x[1,j]))\ .
\end{aligned} \tag{2.6}$$

Combining Eqs. (2.4), (2.5), and (2.6), we get

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

as claimed. ∎

Acknowledgements

A.N. would like to thank Frédéric Magniez for several helpful discussions preceding this work.

References

[1] 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.

[2] 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.

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

[4] 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://http://eccc.hpi-web.de/`, April 18 2010. To appear in FOCS 2010.

[5] 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.

[6] 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.

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

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] 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.

[15] 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.

[16] 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.

[17] 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.

A From streaming algorithms to communication protocols

Here we sketch a proof of Theorem 2.6, highlighting the sole modification we need, namely in the definition of information cost. We refer the reader to [14] 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 $4n^2$. Each such 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 problem f_n defined in Section 2.2. 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 P for ASCENSION(n) with T sequential iterations of messages in the same order as for the one-pass case described in [14, Section 4]. 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 P for ASCENSION(n) gives rise to a protocol for a single instance of f_n through a direct sum property of its "information cost". Let μ_0 be the uniform distribution over the subset of $\left(\{0,1\}^n \times [n] \times \{0,1\}\right)$ on which the function f_n is 0. Let $(\boldsymbol{X}, \boldsymbol{k}, \boldsymbol{c}) = \left(X^i, k^i, c^i\right)_{i=1}^n$ be n instances of f_n, distributed according to μ_0^n. Let R denote the public random bits in the protocol P arising from the randomness used by the streaming algorithm. The information cost of P is defined as:

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

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 $\mathrm{IC}_{\mu_0^n}(P) \leq Ts$, as each message $M_{\mathsf{B}_n,j}$ is of length at most s.

The protocol P may be adapted to n different protocols P'_i, $i \in [n]$, for f_n, by precisely the same method of embedding an instance of f_n into one of ASCENSION(n), as described in [14, Section 4.3]. The $2n$ players in P 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 P does. The information cost of P'_i is measured as

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

where R^i is the public randomness in P'_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

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

as in [14, Lemma 3]. Therefore at least one protocol for f_n from (P'_i), call it P', has information cost at most Ts/n. Note that we may replace Bob's messages by the entire message transcript in P' 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 P', even given Bob's input and the public randomness, is at most sT.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de