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

The space complexity of recognizing well-parenthesized expressions

Rahul Jain* Ashwin Nayak†

April 19, 2010

Abstract

We show an $\Omega(\sqrt{n}/T^3)$ 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.

*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 [4], and is implicit in a myriad of information processing tasks. It has been studied extensively, most recently in setting of streaming algorithms [11]. 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 as possible. (See the text [12] for an introduction to this topic.)

Magniez, Mathieu, and Nayak [11] 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 $\Omega(\sqrt{n}/T^3)$ 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 [11] and establishes the peculiar power of bi-directional streams.

A relatively straightforward generalization of the one-pass algorithm in [11] gives us a unidirectional randomized T-pass streaming algorithm that uses space $O\left(\sqrt{(n \log n)/T}\right)$ and has polynomially small probablity of error. The lower bound we derive thus comes within a factor $\sqrt{\log n}/T^{5/2}$ of optimal. We believe the dependence on T may be improved with a more nuanced analysis. We also note that the bound for one pass algorithms is a factor of $\sqrt{\log n}$ better than the one in [11], 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 by the previous set of authors. 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, [3, 14, 2, 8, 6]), which reduces the task to bounding from below the information cost of a variant of the INDEX problem. It is here that we depart from the earlier route. First, we formulate a bound for protocols resulting from one-pass algorithms entirely in information-theoretic terms. Second, we develop a round reduction type of argument to extend this technique to multi-pass protocols. A notion of information cost for INDEX has been studied previously by Jain, Radhakrishnan, and Sen [7] in the context of privacy in communication. This notion, defined in terms of the hard distribution for the problem, however seems not to be directly relevant to our situation, where we deal with an easy distribution (on which the function value is a constant).

2 Lower bounds for unidirectional streams

In this section we present the main result of this article, 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 between a streaming algorithm for DYCK(2) and two-party communication protocols for a variant of INDEX given by Magniez, Mathieu, and Nayak [11]. 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 bounds in Sections 2.3 and 2.4. 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_1 for bit-strings or integers, and capital letters like X, K, M_1 for random variables over the corresponding sample spaces. We use the same symbol for a random variable and its distribution. Given joint 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 random variable $X \in \{0,1\}^n$, and number $k \in [n]$, we let $X[1,k] = X_1 \cdots X_k$. We use integer subscripts in the standard way to denote elements of a sequence. We denote the ℓ_1-distance between two random variables A, B over the same sample space by $\|A - B\|$. If they are jointly distributed with another random variable C, we denote the ℓ_1-distance between the conditional distributions $A|(C = c)$ and $B|(C = c)$ by $\|(A - B)|(C = c)\|$, or simply by $\|(A - B)|c\|$, if C is clear from the context. When a sample z is drawn from distribution Z, we denote it as $z \leftarrow Z$.

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 [5]. We start with the chain rule for mutual information. Let $\mathrm{I}(X : Y)$ denote the mutual information between two random variables.

Fact 2.1 (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 expectation value of a bounded random variable is continuous as a function of the probability distribution.

Fact 2.2 *Let Z, Z' be two random variables on the same sample space $\mathcal{S}$. Let g be a real-valued function on $\mathcal{S}$ such that $\sup_{s\in\mathcal{S}} |g(s)| \leq G$. Then*

$$|\mathbb{E}_{z\leftarrow Z}\, g(z) - \mathbb{E}_{z\leftarrow Z'}\, g(z)| \quad \leq \quad \left\|Z - Z'\right\| G \ .$$

The following fact bounds the (Shannon) entropy H of a biased Boolean random variable.

Fact 2.3 *Let A be a binary random variable such that $\Pr[A = 0] = \frac{1}{2} + \delta$, for some $\delta \in [-1/2, 1/2]$. Then $1 - \mathrm{H}(A) \geq \frac{2}{\ln 2}\delta^2$.*

The Average encoding theorem [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 *Let KM be joint random variables. Then,*

$$\mathbb{E}_{k\leftarrow K} \|M|k - M\| \quad \leq \quad \sqrt{\kappa\, \mathrm{I}(K : M)} \ ,$$

where κ is the constant $2 \ln 2$.

2.2 The two-party communication problem

We refer the reader to the text [10] for an introduction to the model of two-party communication protocols.

We consider randomized two-party communication protocols arising from streaming algorithms for Dyck languages. 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 [11], and was previously studied in the setting of one-way communication as "serial encoding" [1, 13].

The protocols for f_n on which we focus satisfy further properties that arise from considerations in the streaming model:

1. The protocol consists of an even number $2T$ of messages, beginning with Alice, for some positive integer T. The number T is the number of passes of the streaming algorithm in a single direction; we also call it the number of passes of the communication protocol.
2. Alice's messages are deterministic, i.e., determined entirely by the protocol transcript and her input. Bob has access to private randomness, and his messages may be randomized.
3. Each message from Alice to Bob, or back is of length at most $\ell \geq 0$.
4. The protocol has no memory of messages prior to the one received most recently. Consequently, message m_i sent by a party is computed from the preceding message m_{i-1} (if any), her/his input, and in the case of Bob, also using private randomness.[1]

[1] Our results do not rely on this property of the protocol, but we believe that it may lead to a tighter bounds.

5. Let (X, K) be independent random variables distributed uniformly in $\{0,1\}^n \times [n]$. Let M_i^0 be the ith message, when Alice's input is X and Bob's input is $(K, X[1, K-1], X_K)$, i.e., their inputs are chosen uniformly from the set of 0s of f. Then,

$$\mathrm{I}(K : M_2^0 M_4^0 \cdots M_{2T}^0 | X) \quad \leq \quad cT \ ,$$

for some $c \geq 0$.

6. The distributional error of the protocol under the uniform distribution over inputs x, k, b is at most $\varepsilon < \frac{1}{2}$.

We call protocols as described above $(T, \ell, c, \varepsilon)$ *streaming* protocols for the function f_n.

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

Theorem 2.5 (Magniez, Mathieu, and Nayak [11]) *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 a $(T, s, 2s/n, 2\delta)$ streaming protocol for f_n.*

For completeness, we sketch a proof of this theorem in Appendix A, highlighting the sole difference from the one-pass case.

2.3 The lower bound for one-pass protocols

We explain our method up by showing that one-pass streaming protocols for f_n require messages of length linear in n.

Theorem 2.6 *Any $(1, \ell, c, \varepsilon)$ streaming protocol for f_n satisfies $\sqrt{c} + \sqrt{\ell/n} \geq \frac{1-2\varepsilon}{2\sqrt{2 \ln 2}}$.*

Proof: Consider a $(1, \ell, c, \varepsilon)$ streaming 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.

We show that the random variables with Alice after the second message, conditioned upon the function value being 0 (i.e, when $B = X_K$) are close in ℓ_1 distance to the same variables when the function value is 1 (i.e, when $B = \bar{X}_K$). Let the input and message random variables in the two cases be denoted by

$$\begin{aligned} XKM_1^0M_2^0 &= XK\, M_1(X)\, M_2(X[1,K-1], X_k, M_1(X)), \quad \text{and} \\ XKM_1^1M_2^1 &= XK\, M_1(X)\, M_2(X[1,K-1], \bar{X}_k, M_1(X)) \ , \end{aligned}$$

respectively. We use superscripts $0, 1$ on M_1 for consistency, eventhough $M_1^0(x) = M_1^1(x)$ for every x, irrespective of the function value. The random variables available to Alice at the end of the protocol are X, M_1, M_2. We show below that:

Lemma 2.7 $\left\| XM_1^0M_2^0 - XM_1^1M_2^1 \right\| \quad \leq \quad \left\| XKM_1^0M_2^0 - XKM_1^1M_2^1 \right\| \quad \leq \quad 4\sqrt{\kappa}(\sqrt{c} + \sqrt{\ell/n}\,)$, *where $\kappa = 2 \ln 2$.*

Since the protocol P identifies the two distributions, $XM_1^0M_2^0$ and $XM_1^1M_2^1$, with average error ε, we have $\left\|XM_1^0M_2^0 - XM_1^1M_2^1\right\| \geq 2(1-2\varepsilon)$. The theorem follows. ∎

This immediately gives us a space lower bound for one-pass streaming algorithms for DYCK(2).

Corollary 2.8 *Any randomized one-pass streaming algorithm for* DYCK(2) *that has worst case two-sided error $\delta \leq \frac{1}{4}$ uses space at least*

$$\frac{(1-4\delta)^2}{16(\sqrt{2}+1)^2 \ln 2} \times \sqrt{n} \ .$$

on instances of length n.

Proof of Lemma 2.7: The first inequality follows from the monotonicity of ℓ_1 norm with respect to taking marginals. For the second inequality, we note that since the random variable M_2^0 carries little information about K, they are nearly independent. Define a random variable $\tilde{M}_2$ implicitly by the equation $KX\tilde{M}_2 = K \otimes (XM_2^0)$, where the latter is the product of the two distributions K and the marginal XM_2^0. Then,

Lemma 2.9 $\left\|XK\tilde{M}_2 - XKM_2^0\right\| \quad \leq \quad \sqrt{\kappa \, \mathrm{I}(K : M_2^0|X)}$, *where* $\kappa = 2\ln 2$.

We defer the proof of this lemma to later in the section.

As a corollary of Lemma 2.9, we get that for most (x,k), M_2 does not distinguish between Bob's input being $(x_1,\ldots,x_k)$ or $(x_1,\ldots,x_{k-1})$. We defer the proof to later in this section.

Corollary 2.10 *For all x, define $M_2(x[1,k-2], x_{k-1}, M_1(x))$ as $\tilde{M}_2(x)$ when $k=1$. Then*

$$\mathbb{E}_{(x,k)\leftarrow(X,K)} \left\|M_2(x[1,k-2], x_{k-1}, M_1(x)) - M_2(x[1,k-1], x_k, M_1(x))\right\| \quad \leq \quad 2\sqrt{\kappa c}.$$

We introduce a random variable Y correlated with XKM_1^0. When $XKM_1^0 = (x,k,m_1)$, this random variable is uniformly distributed over the set

$$\left\{y \in \{0,1\}^n \ : \ y[1,k-1] = x[1,k-1], \ y_k = \bar{x}_k, \ M_1(y) = m_1\right\},$$

if the set is non-empty, and is 0^n otherwise.

We claim that the distributions of YKM_1^0 and XKM_1^0 are quite close. The idea is that if M_1^0 is smaller than n in length (if $d = \ell/n < n$), it does not carry much information about many coordinates of X. Therefore, "flipping" the bit of X in a random such coordinate (conditioned on M_1^0) does not perturb the overall distribution much. We formalize this later in this section.

Lemma 2.11 $\|YK - XK\| \quad \leq \quad \left\|YKM_1^0 - XKM_1^0\right\| \quad \leq \quad 2\sqrt{\kappa d}$, *where* $d = \ell/n$.

We now have all the ingredients for the proof of the theorem. Note that the random variable M_2^1 for inputs XK is the same as M_2^0 for inputs YK, as $Y_K = \bar{X}_K$. Since XK and YK are close to each other in distribution, we expect the two random variables M_2^0, M_2^1 to also be close to each other. The complication here is that we would like this to be the case for a random triple (x,k,y) in the support of XKY. It is here that we make essential use of the property that M_2^0 is nearly

independent of k, and therefore nearly the same for k and $k-1$. Since x and y have the same prefix up to the $(k-1)$th coordinate, we conclude that M_2^0 for the two are nearly the same. Formally,

$$\begin{aligned}
\left\|XKM_1^0M_2^0 - XKM_1^1M_2^1\right\| &= \left\|XKM_2^0 - XKM_2^1\right\| \\
&= \mathbb{E}_{(x,k)\leftarrow(X,K)} \left\|M_2(x[1,k-1], x_k, M_1(x)) - M_2(x[1,k-1], \bar{x}_k, M_1(x))\right\| \\
&\leq \mathbb{E}_{(x,k)\leftarrow(X,K)} \left\|M_2(x[1,k-1], x_k, M_1(x)) - M_2(x[1,k-2], x_{k-1}, M_1(x))\right\| \\
&\quad + \mathbb{E}_{(x,k)\leftarrow(X,K)} \left\|M_2(x[1,k-1], \bar{x}_k, M_1(x)) - M_2(x[1,k-2], x_{k-1}, M_1(x))\right\| \ , \\
&\quad \text{by the triangle inequality;} \\
&\leq 2\sqrt{\kappa c} + \mathbb{E}_{(x,k)\leftarrow(X,K)} \left\|M_2(x[1,k-1], \bar{x}_k, M_1(x)) - M_2(x[1,k-2], x_{k-1}, M_1(x))\right\| \ , \\
&\quad \text{by Corollary 2.10;} \\
&= 2\sqrt{\kappa c} + \mathbb{E}_{(x,k,y)\leftarrow(X,K,Y)} \left\|M_2(y[1,k-1], y_k, M_1(y)) - M_2(y[1,k-2], y_{k-1}, M_1(y))\right\| \ , \\
&\quad \text{by the definition of } Y \text{ as a function of } X, K; \\
&= 2\sqrt{\kappa c} + \mathbb{E}_{(y,k)\leftarrow(Y,K)} \left\|M_2(y[1,k-1], y_k, M_1(y)) - M_2(y[1,k-2], y_{k-1}, M_1(y))\right\| \ , \\
&\quad \text{as the expression whose expectation we are taking is a function of } Y, K; \\
&\leq 2\sqrt{\kappa c} + 2\left\|XK - YK\right\| \\
&\quad + \mathbb{E}_{(x,k)\leftarrow(X,K)} \left\|M_2(x[1,k-1], x_k, M_1(x)) - M_2(x[1,k-2], x_{k-1}, M_1(x))\right\| \ , \\
&\quad \text{by Fact 2.2;} \\
&\leq 2\sqrt{\kappa c} + 4\sqrt{\kappa d} + 2\sqrt{\kappa c} = 4\sqrt{\kappa}(\sqrt{c} + \sqrt{d}) \ , \\
&\quad \text{by Lemma 2.11 and Corollary 2.10,}
\end{aligned}$$

which concludes the proof. ■

We return to the deferred proofs.

Proof of Lemma 2.9: From the average encoding theorem, Fact 2.4, we have that for every $x \in \{0,1\}^n$,

$$\mathbb{E}_{k\leftarrow K} \left\| M_2(x[1,k-1], x_k, M_1(x)) - \tilde{M}_2(x) \right\| \leq \left[\kappa \, \mathrm{I}(K : M_2^0 | X = x)\right]^{1/2} .$$

By the Jensen inequality,

$$\mathbb{E}_{(x,k)\leftarrow(X,K)} \left\| M_2(x[1,k-1], x_k, M_1(x)) - \tilde{M}_2(x) \right\| \leq \left[\kappa \, \mathrm{I}(K : M_2^0 | X)\right]^{1/2} .$$

The LHS of the above equation is the same as that in the statement of the lemma. ■

Proof of Corollary 2.10: By the triangle inequality,

$$\begin{aligned}
&\mathbb{E}_{(x,k)\leftarrow(X,K)} \|M_2(x[1,k-2], x_{k-1}, M_1(x)) - M_2(x[1,k-1], x_k, M_1(x))\| \\
&\quad\leq \mathbb{E}_{(x,k)\leftarrow(X,K)} \left\|M_2(x[1,k-2], x_{k-1}, M_1(x)) - \tilde{M}_2(x)\right\| \\
&\qquad + \mathbb{E}_{(x,k)\leftarrow(X,K)} \left\|\tilde{M}_2(x) - M_2(x[1,k-1], x_k, M_1(x))\right\| \\
&\quad\leq 2\sqrt{\kappa\, \mathrm{I}(K : M_2^0|X)} \quad\leq\quad 2\sqrt{\kappa c}\ ,
\end{aligned}$$

by the hypothesis in Theorem 2.6 that $\mathrm{I}(K : M_2^0|X) \leq c$. ■

Proof of Lemma 2.11: For all $k \in [n]$ and values $(x[1,k-1], m_1)$ taken by the random variables $X[1,K-1], M_1$, let

$$\delta\big(X_k|(x[1,k-1],m_1)\big) \quad = \quad \Pr[X_k = 0|(x[1,k-1],m_1)] - \frac{1}{2}\ ,$$

where conditioning on $(x[1,k-1],m_1)$ is shorthand for conditioning on the event $(X[1,k-1],M_1) = (x[1,k-1],m_1)$. We have:

Lemma 2.12 *For all k and $(x[1,k-1],m_1)$,*

$$\|(Y-X)|(x[1,k-1],m_1)\| \quad = \quad 4\left|\delta\big(X_k|(x[1,k-1],m_1)\big)\right|\ .$$

Proof: Fix k and $(x[1,k-1],m_1)$. Let $\delta = \delta\big(X_k|(x[1,k-1],m_1)\big)$,

$$\begin{aligned}
Q_0 &= X|(x[1,k-1],m_1,X_k=0), \quad \text{and} \\
Q_1 &= X|(x[1,k-1],m_1,X_k=1)\ .
\end{aligned}$$

Observe that

$$\begin{aligned}
X|(x[1,k-1],m_1) &= \left(\frac{1}{2}+\delta\right)Q_0 + \left(\frac{1}{2}-\delta\right)Q_1, \quad \text{and} \\
Y|(x[1,k-1],m_1) &= \left(\frac{1}{2}+\delta\right)Q_1 + \left(\frac{1}{2}-\delta\right)Q_0\ .
\end{aligned}$$

Therefore

$$\|(Y|(x[1,k-1],m_1) - X|(x[1,k-1],m_1)\| \quad = \quad 2\,|\delta|\cdot\|Q_0-Q_1\| \quad = \quad 4|\delta|\ ,$$

where equality holds because Q_0 and Q_1 have disjoint support. ■

Since $|M_1^0| \leq \ell$, by repeated application of the chain rule for mutual information, Fact 2.1,

$$\ell \quad\geq\quad \mathrm{I}(X : M_1^0) \quad = \quad n\cdot\mathbb{E}_{k\leftarrow K}\mathrm{I}(X_k : M_1^0|X[1,k-1])\ . \tag{2.1}$$

Moreover,

$$\begin{aligned}
&\mathbb{E}_{k\leftarrow K}\mathrm{I}(X_k : M_1^0|X[1,k-1]) \\
&\quad = \mathbb{E}_{(k,m_1)\leftarrow(KM_1^0)}\big(1-\mathrm{H}(X_k|X[1,k-1],m_1)\big) \\
&\quad = \mathbb{E}_{(k,m_1,x[1,k-1])\leftarrow(KM_1^0X[1,k-1])}\big(1-\mathrm{H}(X_k|x[1,k-1],m_1)\big) \\
&\quad \geq \frac{2}{\ln 2}\mathbb{E}_{(k,m_1,x[1,k-1])\leftarrow(KM_1^0X[1,k-1])}\,\delta\big(X_k|(x[1,k-1],m_1)\big)^2, \\
&\qquad \text{by Fact 2.3;} \\
&\quad \geq \frac{2}{\ln 2}\left(\mathbb{E}_{(k,m_1,x[1,k-1])\leftarrow(KM_1^0X[1,k-1])}\left|\delta\big(X_k|(x[1,k-1],m_1)\big)\right|\right)^2, \\
&\qquad \text{by the Jensen inequality;} \\
&\quad = \frac{2}{\ln 2}\left(\frac{1}{4}\cdot\mathbb{E}_{(k,m_1,x[1,k-1])\leftarrow(KM_1^0X[1,k-1])}\left\|(Y|(x[1,k-1],m_1)-X|(x[1,k-1],m_1)\right\|\right)^2, \\
&\qquad \text{by Lemma 2.12;} \\
&\quad = \frac{1}{8\ln 2}\left\|YKM_1^0-XKM_1^0\right\|^2 \ .
\end{aligned}$$

Along with Eq. (2.1), this establishes the second inequality in the statement of the lemma; the first inequality follows from monotonicity of the ℓ_1 norm with respect to taking marginals. ■

2.4 The lower bound for T-pass protocols

Using the technique developed for one-pass protocols, we now derive a bound for multi-pass streaming protocols for f_n.

Theorem 2.13 *Any (T,ℓ,c,ε) streaming protocol for f_n satisfies*

$$2\sqrt{c}+\sqrt{\frac{\ell}{n}} \quad \geq \quad \frac{(1-2\varepsilon)}{\sqrt{2\kappa}\,T^{3/2}} \ ,$$

where $\kappa = 2\ln 2$.

Proof: Consider a (T,ℓ,c,ε) streaming 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.

We compare the random variables in Alice's possession at the end of T rounds, when the function value is 0 and when it is 1. As before, we use superscripts 0 and 1 to mark messages that are sent by the two parties when $B = X_K$ or when $B = \bar{X}_K$, respectively. In particular, for every $t \in [T]$, $XKM_1^0M_2^0\cdots M_{2t}^0$ denote the jointly distributed random variables corresponding to the the inputs and the first $2t$ messages in the protocol P, when $B = X_K$. Similarly, $XKM_1^1M_2^1\cdots M_{2t}^1$ denote the analogous random variables when $B = \bar{X}_K$.

Consider the random variables $XM_{2t-1}M_{2t}$ available to Alice at the end of t passes, $t \in [T]$. We prove that Alice's ability to distinguish between the two values of f_n from these random variables does not increase by much in any pass. In fact, we prove something stronger:

Lemma 2.14 *For every $t \in [T]$,*

$$\begin{aligned}&\left\|XKM_1^0\cdots M_{2t}^0 - XKM_1^1\cdots M_{2t}^1\right\| \\ &\quad\leq \left\|XKM_1^0\cdots M_{2t-2}^0 - XKM_1^1\cdots M_{2t-2}^1\right\| + 2\sqrt{2\kappa T}\left(\sqrt{d}+2\sqrt{c}\right) \ ,\end{aligned}$$

where $d = \ell/n$, and $\kappa = 2\ln 2$.

The proof of this lemma is presented later in this section.

Repeatedly applying Lemma 2.14, we conclude that Alice's ability to distinguish between the function values 0 and 1 from the random variables in her possession is limited:

$$\begin{aligned}&\left\|XM_{2T-1}^0M_{2T}^0 - XM_{2T-1}^1M_{2T}^1\right\| \\ &\quad\leq \left\|XKM_1^0\cdots M_{2T}^0 - XKM_1^1\cdots M_{2T}^1\right\| \\ &\quad\leq 2\sqrt{2\kappa}(2\sqrt{c}+\sqrt{d}\,)\,T^{3/2} \ . \qquad (2.2)\end{aligned}$$

Since the protocol P computes f_n with average error ε, Alice can identify the two distributions $XM_{2T-1}^0M_{2T}^0$ and $XM_{2T-1}^1M_{2T}^1$ with average error ε. Hence, we have

$$\left\|XM_1^0M_2^0\cdots M_{2T}^0 - XM_1^1M_2^1\cdots M_{2T}^1\right\| \quad\geq\quad 2(1-2\varepsilon) \ . \qquad (2.3)$$

The theorem follows by combining Eq. (2.2) and (2.3). ∎

Along with Theorem 2.5, this gives us a space lower bound for T-pass streaming algorithms for Dyck(2).

Corollary 2.15 *Any randomized unidirectional T-pass streaming algorithm for* Dyck(2) *that has worst case two-sided error $\delta \leq \frac{1}{4}$ uses space at least*

$$\frac{(1-4\delta)^2}{8(2\sqrt{2}+1)^2\ln 2}\times\frac{\sqrt{n}}{T^3} \ .$$

on instances of length n.

We build up to the proof of Lemma 2.14, starting with analogues of Lemma 2.9 and Corollary 2.10. Define the random variables $\tilde{M}_i$, for $i \in [2T]$ implicitly by the equation

$$KX\tilde{M}_1\tilde{M}_2\cdots\tilde{M}_{2T} \quad = \quad K\otimes(XM_1^0\cdots M_{2T}^0).$$

Fix $t \in [T]$. The first lemma states that we may think of K and $M_1^0\cdots M_{2t-1}^0$ as being independent, conditioned on X, at a small cost.

Lemma 2.16 *For all $t \in [T]$,*

$$\left\|KXM_1^0\cdots M_{2t-1}^0 - KX\tilde{M}_1\cdots\tilde{M}_{2t-1}\right\| \quad\leq\quad \left[\kappa\,\mathrm{I}(M_2^0M_4^0\cdots M_{2t-2}^0 : K|X)\right]^{1/2}.$$

The essence of the second lemma is that on average, given all previous messages, the message M_{2t} does not distinguish heavily between K and $K-1$.

Lemma 2.17 *For all $t \in [T]$,*

$$\begin{aligned}\mathbb{E}_{(x,k,m_1,\ldots,m_{2t-1}) \leftarrow XKM_1^0 \cdots M_{2t-1}^0} & \|M_{2t}(x[1,k-1], x_k, m_{2t-1}) - M_{2t}(x[1,k-2], x_{k-1}, m_{2t-1})\| \\ & \leq \quad 2\big[\kappa \, \mathrm{I}(M_{2T}^0 : K | X M_2^0 M_4^0 \cdots M_{2t-2}^0)\big]^{1/2},\end{aligned}$$

where we interpret $M_{2t}(x[1,k-2], x_{k-1}, m_{2t-1})$ as $\tilde{M}_{2t}(x, m_{2t-1})$ when $k = 1$.

The proofs of the two lemmas rely on the fact that Alice's messages (i.e., M_i for odd i) are deterministic. They are much the same as the proofs for one-pass protocols, and are omitted.

For every $t \in [T]$, define a random variable Y_t correlated with $XKM_1^0 \cdots M_{2t-1}^0$ as follows. For any fixed value $(x, k, m_1, \ldots, m_{2t-1})$ of these random variables, let Y_t be uniformly distributed in the set

$$\big\{y \in \{0,1\}^n \ : \ y[1,k-1] = x[1,k-1], y_k = \bar{x}_k, \text{ and } \\ M_1(y) = m_1, M_{2j-1}(y, m_{2j-2}) = m_{2j-1} \text{ for } 2 \leq j \leq t\big\},$$

if the set is non-empty, and $Y_t = 0^n$ otherwise. For all $k \in [n]$ and values $(x[1,k-1], m_1, \ldots, m_{2t-1})$ for $X[1,k-1], M_1^0 M_2^0 \cdots M_{2t-1}^0$, define

$$\begin{aligned}& \delta\big(X_k | (x[1,k-1], m_1, \ldots, m_{2t-1})\big) \\ & \quad = \quad \Pr\big[X_k = 0 \mid (x[1,k-1], m_1, \ldots, m_{2t-1}))\big] - \frac{1}{2} \ .\end{aligned}$$

As before,

Lemma 2.18 *For all $t \in [T]$, $k \in [n]$, and values $(x[1,k-1], m_1, \ldots, m_{2t-1})$ for the random variables $(X[1,k-1], M_1^0 M_2^0 \cdots M_{2t-1}^0)$,*

$$\|(Y_t - X) | (x[1,k-1], m_1, \ldots, m_{2t-1})\| \quad = \quad 4\left|\delta\big(X_k | (x[1,k-1], m_1, \ldots, m_{2t-1})\big)\right| \ .$$

The proof is similar to that of Lemma 2.12, and is omitted.

If the messages from Alice are short, and Bob's messages do not contain information about the index k, the protocol is unable to distinguish between X and Y within the first $2t-1$ messages. A crucial step in showing this is the use of $\tilde{M}_1 \tilde{M}_2 \cdots \tilde{M}_{2t-1}$ to masquerade as the actual sequence of messages. This may be viewed as a kind of round elimination. At a technical level, it is required to make the messages independent of the index k, which allows us to use the chain rule for mutual information.

Lemma 2.19 *For all $t \in [T]$,*

$$\left\|XKM_1^0 \cdots M_{2t-1}^0 - Y_t K M_1^0 \cdots M_{2t-1}^0\right\| \quad \leq \quad 2\sqrt{2\kappa t d} + 2\left[\kappa \, \mathrm{I}(M_2^0 M_4^0 \cdots M_{2t-2}^0 : K | X)\right]^{1/2},$$

where $d = \ell / n$.

Proof: We have

$$
\begin{aligned}
&\left\|XKM_1^0\cdots M_{2t-1}^0 - YKM_1^0\cdots M_{2t-1}^0\right\| \\
&= \mathbb{E}_{(k,x[1,k-1],m_1,\ldots,m_{2t-1})\leftarrow(KX[1,K-1]M_1^0\cdots M_{2t-1}^0)} \left\|(X-Y)|(k,x[1,k-1],m_1,\ldots,m_{2t-1})\right\| \\
&= 4\,\mathbb{E}_{(k,x[1,k-1],m_1,\ldots,m_{2t-1})\leftarrow(KX[1,K-1]M_1^0\cdots M_{2t-1}^0)} \left|\delta\big(X_k|(x[1,k-1],m_1,\ldots,m_{2t-1})\big)\right| \\
&\quad \text{By Lemma 2.18;} \\
&\\
&\leq 4\cdot\mathbb{E}_{(k,x[1,k-1],m_1,\ldots,m_{2t-1})\leftarrow(KX[1,K-1]\tilde{M}_1\cdots\tilde{M}_{2t-1})} \left|\delta\big(X_k|(x[1,k-1],m_1,\ldots,m_{2t-1})\big)\right| \\
&\quad + 2\left\|KX[1,K-1]M_1^0\cdots M_{2t-1}^0 - KX[1,K-1]\tilde{M}_1\cdots\tilde{M}_{2t-1}\right\|, \\
&\quad \text{by Fact 2.2;} \\
&\\
&\leq 4\,\mathbb{E}_{(k,x[1,k-1],m_1,\ldots,m_{2t-1})\leftarrow(KX[1,K-1]\tilde{M}_1\cdots\tilde{M}_{2t-1})} \left|\delta\big(X_k|(x[1,k-1],m_1,\ldots,m_{2t-1})\big)\right| \\
&\quad + 2\left[\kappa\,\mathrm{I}(M_2^0M_4^0\cdots M_{2t-2}^0 : K|X)\right]^{1/2}, \\
&\quad \text{by Lemma 2.16;} \\
&\\
&\leq 2\cdot\mathbb{E}_{(k,x[1,k-1],m_1,\ldots,m_{2t-1})\leftarrow(KX[1,K-1]\tilde{M}_1\cdots\tilde{M}_{2t-1})} \left[\kappa\big(1-\mathrm{H}(X_k|(x[1,k-1],m_1,\ldots,m_{2t-1}))\big)\right]^{1/2} \\
&\quad + 2\left[\kappa\,\mathrm{I}(M_2^0M_4^0\cdots M_{2t-2}^0 : K|X)\right]^{1/2}, \\
&\quad \text{by Fact 2.3;} \\
&\\
&\leq 2\left[\kappa\,\mathbb{E}_{(k,x[1,k-1],m_1,\ldots,m_{2t-1})\leftarrow(KX[1,K-1]\tilde{M}_1\cdots\tilde{M}_{2t-1})}\big(1-\mathrm{H}(X_k|(x[1,k-1],m_1,\ldots,m_{2t-1}))\big)\right]^{1/2} \\
&\quad + 2\left[\kappa\,\mathrm{I}(M_2^0M_4^0\cdots M_{2t-2}^0 : K|X)\right]^{1/2}, \\
&\quad \text{by the Jensen inequality;} \\
&\\
&= 2\left[\kappa\,\mathbb{E}_{k\leftarrow K}\mathrm{I}(X_k : \tilde{M}_1\cdots\tilde{M}_{2t-1}|X[1,k-1])\right]^{1/2} + 2\left[\kappa\,\mathrm{I}(M_2^0M_4^0\cdots M_{2t-2}^0 : K|X)\right]^{1/2} \\
&= 2\left[(\kappa/n)\,\mathrm{I}(X : \tilde{M}_1\cdots\tilde{M}_{2t-1})\right]^{1/2} + 2\left[\kappa\,\mathrm{I}(M_2^0M_4^0\cdots M_{2t-2}^0 : K|X)\right]^{1/2} \\
&\quad \text{by the chain rule, Fact 2.1;} \\
&\\
&\leq 2\sqrt{2\kappa t\ell/n} + 2\left[\kappa\,\mathrm{I}(M_2^0M_4^0\cdots M_{2t-2}^0 : K|X)\right]^{1/2}, \\
&\quad \text{since for every } i, |\tilde{M}_i| \leq \ell.
\end{aligned}
$$

■

Finally, we turn to the main lemma. For this, consider "mixed" runs of the protocol P in which B is set to X_K in the first $t-1$ passes, and to $\bar{X}_K$ in the t'th pass. The associated message random variables in these runs are written as $XKM_1^0M_2^0\cdots M_{2t-1}^0M_{2t}^1$.

Proof of Lemma 2.14: Let

$$\alpha \quad = \quad \left\|XKM_1^1\cdots M_{2t-1}^1 - XKM_1^0\cdots M_{2t-1}^0\right\|,$$

which is also equal to $\left\|XKM_1^1\cdots M_{2t-2}^1 - XKM_1^0\cdots M_{2t-2}^0\right\|$. The first two inequalities below follow from the triangle inequality. The rest are explained as we derive them.

$$
\begin{aligned}
&\left\|XKM_1^0\cdots M_{2t}^0 - XKM_1^1\cdots M_{2t}^1\right\| \\
&\le \left\|XKM_1^1\cdots M_{2t}^1 - XKM_1^0\cdots M_{2t-1}^0M_{2t}^1\right\| + \left\|XKM_1^0\cdots M_{2t}^0 - XKM_1^0\cdots M_{2t-1}^0M_{2t}^1\right\| \\
&= \left\|XKM_1^1\cdots M_{2t-1}^1 - XKM_1^0\cdots M_{2t-1}^0\right\| \\
&\quad + \mathbb{E}_{(x,k,m_1,\ldots,m_{2t-1})\leftarrow XKM_1^0\cdots M_{2t-1}^0} \left\|M_{2t}(x[1,k-1],x_k,m_{2t-1}) - M_{2t}(x[1,k-1],\bar{x}_k,m_{2t-1})\right\| \\
&\le \alpha + \mathbb{E}_{(x,k,m_1,\ldots,m_{2t-1})\leftarrow XKM_1^0\cdots M_{2t-1}^0} \left\|M_{2t}(x[1,k-1],x_k,m_{2t-1}) - M_{2t}(x[1,k-2],x_{k-1},m_{2t-1})\right\| \\
&\quad + \mathbb{E}_{(x,k,m_1,\ldots,m_{2t-1})\leftarrow XKM_1^0\cdots M_{2t-1}^0} \left\|M_{2t}(x[1,k-2],x_{k-1},m_{2t-1}) - M_{2t}(x[1,k-1],\bar{x}_k,m_{2t-1})\right\| \\
&\le \alpha + 2\left[\kappa\ \mathrm{I}(M_{2t}^0 : K|XM_2^0M_4^0\cdots M_{2t-2}^0)\right]^{1/2} \\
&\quad + \mathbb{E}_{(x,k,m_1,\ldots,m_{2t-1})\leftarrow XKM_1^0\cdots M_{2t-1}^0} \left\|M_{2t}(x[1,k-2],x_{k-1},m_{2t-1}) - M_{2t}(x[1,k-1],\bar{x}_k,m_{2t-1})\right\|, \\
&\quad \text{by Lemma 2.17;} \\
&= \alpha + 2\left[\kappa\ \mathrm{I}(M_{2t}^0 : K|XM_2^0M_4^0\cdots M_{2t-2}^0)\right]^{1/2} \\
&\quad + \mathbb{E}_{(y,x,k,m_1,\ldots,m_{2t-1})\leftarrow YXKM_1^0\cdots M_{2t-1}^0} \left\|M_{2t}(y[1,k-2],y_{k-1},m_{2t-1}) - M_{2t}(y[1,k-1],y_k,m_{2t-1})\right\| \\
&\quad \text{by definition of } Y; \\
&= \alpha + 2\left[\kappa\ \mathrm{I}(M_{2t}^0 : K|XM_2^0M_4^0\cdots M_{2t-2}^0)\right]^{1/2} \\
&\quad + \mathbb{E}_{(y,k,m_1,\ldots,m_{2t-1})\leftarrow YKM_1^0\cdots M_{2t-1}^0} \left\|M_{2t}(y[1,k-2],y_{k-1},m_{2t-1}) - M_{2t}(y[1,k-1],y_k,m_{2t-1})\right\| \\
&\le \alpha + 2\left[\kappa\ \mathrm{I}(M_{2t}^0 : K|XM_2^0M_4^0\cdots M_{2t-2}^0)\right]^{1/2} \\
&\quad + \mathbb{E}_{(x,k,m_1,\ldots,m_{2t-1})\leftarrow XKM_1^0\cdots M_{2t-1}^0} \left\|M_{2t}(x[1,k-2],x_{k-1},m_{2t-1}) - M_{2t}(x[1,k-1],x_k,m_{2t-1})\right\| \\
&\quad + \left\|XKM_1^0\cdots M_{2t-1}^0 - YKM_1^0\cdots M_{2t-1}^0\right\| \\
&\quad \text{by Fact 2.2;} \\
&\le \alpha + 4\left[\kappa\ \mathrm{I}(M_{2t}^0 : K|XM_2^0\cdots M_{2t-2}^0)\right]^{1/2} + \left\|XKM_1^0\cdots M_{2t-1}^0 - YKM_1^0\cdots M_{2t-1}^0\right\| \\
&\quad \text{by Lemma 2.17;} \\
&\le \alpha + 4\left[\kappa\ \mathrm{I}(M_{2t}^0 : K|XM_2^0M_4^0\cdots M_{2t-2}^0)\right]^{1/2} \\
&\quad + 2\sqrt{2\kappa td} + 2\left[\kappa\ \mathrm{I}(M_2^0M_4^0\cdots M_{2t-2}^0 : K|X)\right]^{1/2} \\
&\quad \text{by Lemma 2.19;} \\
&\le \alpha + 2\sqrt{2\kappa td} + 4\sqrt{2\kappa}\left[\mathrm{I}(M_{2t}^0 : K|XM_2^0M_4^0\cdots M_{2t-2}^0) + I(M_2^0M_4^0\cdots M_{2t-2}^0 : K|X)\right]^{1/2},
\end{aligned}
$$

by the Jensen inequality. Applying the chain rule, Fact 2.1, we get

$$\begin{aligned}
&\left\|XKM_1^0\cdots M_{2t}^0 - XKM_1^1\cdots M_{2t}^1\right\| \\
&\leq \alpha + 2\sqrt{2\kappa td} + 4\sqrt{2\kappa}\left[\mathrm{I}(M_2^0M_4^0\cdots M_{2t}^0 : K|X)\right]^{1/2} \\
&\leq \alpha + 2\sqrt{2\kappa td} + 4\sqrt{2\kappa Tc}
\end{aligned}$$

since by hypothesis, $\mathrm{I}(M_2^0M_4^0\cdots M_{2t}^0 : K|X) \leq Tc$. This gives us the claimed bound. ∎

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] Z. Bar-Yossef, T. S. Jayram, R. 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] A. Chakrabarti, Y. Shi, A. Wirth, and A. 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.

[4] 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.

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

[6] R. Jain, J. Radhakrishnan, and P. 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.

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

[8] 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.

[9] 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.

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

[11] 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*, New York, NY, June 6–8 2010. ACM Press. To appear.

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

[13] 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.

[14] M. Saks and X. 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 2.5, highlighting the sole modification we need, namely in the definition of information cost. We refer the reader to [11] 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 A_i, 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 [11, 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 $M_{B_n,j}$, $j \in [T]$, 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 $\big(\{0,1\}^n \times [n] \times \{0,1\}\big)$ 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 P arising from the randomness used by the streaming algorithm. The information cost of P is defined as:

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

This is the natural and straightforward extension of the measure used in the one-pass case, which concentrates on $M_{B_n,1}$, the single message sent by B_n. Note that $\mathrm{IC}_{\mu_0^n}(P) \leq Ts$, as each message $M_{B_n,j}$ 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 [11, Section 4.3]. The $2n$ players in P are simulated by two players, Alice and Bob, as before: Alice simulates $A_1, B_1, A_2, B_2, \ldots, A_i$, sends a message to Bob, who simulates $B_i, A_{i+1}, B_{i+1}, \ldots, A_n, B_n$, sends a message to Alice, who simulates $A_n, A_{n-1}, \ldots, A_1$, 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

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

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 [11, Lemma 3]. Therefore at least one protocol for f_n from (P'_i), call it P', has information cost at most Ts/n. Finally, as in [11, Lemma 1], we may make Alice deterministic in P', at the cost of increasing the distributional error (under the uniform distribution over inputs) to at most 2δ, and the information cost to $2Ts/n$.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de