

# Realizable Paths and the NL vs L Problem

Shiva Kintali

College of Computing,  
Georgia Institute of Technology,  
Atlanta, GA 30332-0765.  
kintali@cc.gatech.edu

## Abstract

A celebrated theorem of Savitch [Sav70] states that  $NSPACE(S) \subseteq DSPACE(S^2)$ . In particular, Savitch gave a deterministic algorithm to solve ST-CONNECTIVITY (an **NL**-complete problem) using  $O(\log^2 n)$  space, implying  $NL \subseteq DSPACE(\log^2 n)$ . While Savitch's theorem itself has not been improved in the last four decades, studying the space complexity of several special cases of ST-CONNECTIVITY has provided new insights into the space-bounded complexity classes.

In this paper, we introduce new kind of graph connectivity problems which we call *graph realizability problems*. All of our graph realizability problems are generalizations of UNDIRECTED ST-CONNECTIVITY. ST-REALIZABILITY, the most general graph realizability problem, is **LogCFL**-complete. We define the corresponding complexity classes that lie between **L** and **LogCFL** and study their relationships.

As special cases of our graph realizability problems we define two natural problems, BALANCED ST-CONNECTIVITY and POSITIVE BALANCED ST-CONNECTIVITY, that lie between **L** and **NL**. We present a deterministic  $O(\log n \log \log n)$  space algorithm for BALANCED ST-CONNECTIVITY. More generally we prove that **SGSLogCFL**, a generalization of BALANCED ST-CONNECTIVITY, is contained in  $DSPACE(\log n \log \log n)$ . To achieve this goal we generalize several concepts (such as graph squaring and transitive closure) and algorithms (such as parallel algorithms) known in the context of UNDIRECTED ST-CONNECTIVITY.

**Keywords:** auxiliary pushdown automata, **LogCFL**, parallel graph algorithms, Savitch's theorem, space-bounded computation, st-connectivity, symmetric Turing machines.

# 1 Introduction

A celebrated theorem of Savitch [Sav70] states that  $NSPACE(S) \subseteq DSPACE(S^2)$ . In particular, Savitch gave a deterministic algorithm to solve ST-CONNECTIVITY (an **NL**-complete problem) using  $O(\log^2 n)$  space, implying  $\mathbf{NL} \subseteq DSPACE(\log^2 n)$ . Savitch's algorithm runs in time  $2^{O(\log^2 n)}$ . It has been a long-standing open problem to improve Savitch's theorem i.e., to prove (i)  $\mathbf{NL} \subseteq DSPACE(o(\log^2 n))$  or (ii)  $\mathbf{NL} \subseteq \mathbf{SC}^2$ , i.e., ST-CONNECTIVITY can be solved by a deterministic algorithm in polynomial time and  $O(\log^2 n)$  space.

While Savitch's theorem itself has not been improved in the last four decades, studying the space complexity of several special cases of ST-CONNECTIVITY has provided new insights into the space-bounded complexity classes. Allender's survey [All07] gives an update of progress related to several special cases of ST-CONNECTIVITY. Recently ST-CONNECTIVITY in planar DAGs with  $O(\log n)$  sources is shown to be in **L** [SBV10]. Stolee and Vinodchandran proved that ST-CONNECTIVITY in DAGs with  $2^{O(\sqrt{\log n})}$  sources embedded on surfaces of genus  $2^{O(\sqrt{\log n})}$  is in **L** [SV10].

All the connectivity problems considered in the literature so far are essentially special cases of ST-CONNECTIVITY. In the first half of this paper, we introduce new kind of graph connectivity problems which we call *graph realizability problems*. All of our graph realizability problems are generalizations of UNDIRECTED ST-CONNECTIVITY. ST-REALIZABILITY, the most general graph realizability problem is **LogCFL**-complete. We define the corresponding complexity classes that lie between **L** and **LogCFL** and study their relationships. As special cases of our graph realizability problems we define two natural problems, BALANCED ST-CONNECTIVITY and POSITIVE BALANCED ST-CONNECTIVITY, that lie between **L** and **NL**.

In the second half of this paper, we study the space complexity of **SGSLogCFL** (see Section 4.1 for definition). We define generalizations of graph squaring and transitive closure, present efficient parallel algorithms for **SGSLogCFL** and use the techniques of Trifonov [Tri08] to show that **SGSLogCFL** is contained in  $DSPACE(\log n \log \log n)$ . This implies that BALANCED ST-CONNECTIVITY, a natural graph connectivity problem which lies between **L** and **NL**, is contained in  $DSPACE(\log n \log \log n)$ .

## 1.1 Preliminaries, Related Work and Our Results

**Auxiliary Pushdown Automata** : A language is accepted by a non-deterministic pushdown automaton (PDA) if and only if it is a context-free language. Deterministic context-free languages are those accepted by the deterministic PDAs. **LogCFL** is the set of all languages that are log-space reducible to a context-free language. Similarly, **LogDCFL** is the set of all languages that are log-space reducible to a deterministic context-free language. There are many equivalent characterizations of **LogCFL**. Sudborough [Sud78] gave the machine class equivalence. Ruzzo [Ruz80] gave an alternating Turing machine (ATM) class equivalent to **LogCFL**. Venkateswaran [Ven91] gave a circuit characterization and showed that  $\mathbf{LogCFL} = \mathbf{SAC}^1$ . For a survey of parallel complexity classes and **LogCFL** see Limaye's thesis [Lim05].

An Auxiliary Pushdown Automaton (NAuxPDA or simply AuxPDA), introduced by Cook [Coo71], is a two-way pushdown automaton augmented with a  $S(n)$ -space bounded working tape. If a deterministic two-way PDA is augmented with a  $S(n)$ -space bounded working tape then we get a Deterministic Auxiliary Pushdown Automaton (DAuxPDA). We present the formal definitions in the **appendix** (see Section A). Let  $NAuxPDA\text{-SpaceTime}(S(n), T(n))$  be the class of languages accepted by an AuxPDA with  $S(n)$ -space bounded working tapes and the running time bounded by  $T(n)$ . Let the corresponding deterministic class be  $DAuxPDA\text{-SpaceTime}(S(n), T(n))$ . It is easy to see that  $\mathbf{NL} \subseteq NAuxPDA\text{-SpaceTime}(O(\log n), poly(n))$ . It is shown by Sudborough that  $NAuxPDA\text{-SpaceTime}(O(\log n), poly(n)) = \mathbf{LogCFL}$

and  $DAuxPDA\text{-SpaceTime}(O(\log n), poly(n)) = \mathbf{LogDCFL}$  [Sud78]. Using ATM simulations, Ruzzo showed that  $\mathbf{LogCFL} \subseteq \mathbf{NC}^2$  [Ruz80]. Simpler proofs of  $DAuxPDA\text{-SpaceTime}(O(\log n), poly(n)) = \mathbf{LogDCFL}$  and  $\mathbf{LogCFL} = \mathbf{SAC}^1$  are given in [MRV99].

Many proof techniques and results obtained in the context of  $\mathbf{NL}$ , are generalized to obtain the corresponding results for  $\mathbf{LogCFL}$ . For example : (i) Borodin [Bor77] proved that  $\mathbf{NL} \subseteq \mathbf{NC}^2$ . Ruzzo [Ruz80] introduced tree-size-bounded alternating Turing machines, gave a new characterization of  $\mathbf{LogCFL}$ , and proved that  $\mathbf{LogCFL} \subseteq \mathbf{NC}^2$ . (ii) Immerman [Imm88] and Szelepcsényi [Sze87] proved that  $\mathbf{NL} = \mathbf{co-NL}$ . Borodin et. al. [BCD<sup>+</sup>89] generalized their inductive counting technique and proved that  $\mathbf{LogCFL} = \mathbf{co-LogCFL}$ . In fact, they proved a stronger result showing that  $\mathbf{SAC}^i$  is closed under complementation for  $i > 0$ . (iii) Wigderson [Wig94] proved that  $\mathbf{NL} \leq_r \oplus \mathbf{NL}$ . Gál and Wigderson [GW96] proved that  $\mathbf{LogCFL} \leq_r \oplus \mathbf{LogCFL}$ . (iv) Nisan [Nis94] proved that  $\mathbf{BPL} \subseteq \mathbf{SC}^2$ . Venkateswaran [Ven06] proved that  $\mathbf{BPLogCFL} \subseteq \mathbf{SC}^2$  and  $\mathbf{BPLogCFL} \subseteq \mathbf{NC}^2$ . Here  $\mathbf{BPLogCFL}$  (resp.  $\mathbf{RLogCFL}$  and  $\mathbf{ZPLogCFL}$ ) is the bounded error (resp. one-sided error and zero error) probabilistic version of  $\mathbf{LogCFL}$ . All the above results are elegant and non-trivial generalizations of the corresponding results in the logspace setting.

Throughout this paper, we consider  $O(\log n)$ -space bounded and polynomial-time bounded AuxPDAs. The *surface configuration* (introduced by Cook [Coo71]) of an AuxPDA, on an input  $w$ , consists of the state, contents and head positions of the work tapes, the head position of the input tape and the topmost symbol of the stack i.e., the rightmost symbol of the pushdown tape. Note that for a space  $S(n)$ -bounded AuxPDA, its surface configurations take only  $O(S(n))$  space. In the rest of the paper, we will refer to surface configurations as configurations. For an input  $w$ , a pair of configurations  $(C_1, C_2)$  is *realizable* if the AuxPDA can move from  $C_1$  to  $C_2$  ending with its stack at the same height as in  $C_1$ , and without popping its stack below its level in  $C_2$  for any of the intermediate configurations. An AuxPDA  $M$  accepts an input  $w$  iff there is a realizable pair  $(I, A)$ , where  $I$  is the initial configuration and  $A$  is the unique accepting configuration.

**Realizable Paths** :  $\mathbf{ST-CONNECTIVITY}$  (resp.  $\mathbf{UNDIRECTED ST-CONNECTIVITY}$ ) is the problem of determining whether there exists a path between two distinguished vertices  $s$  and  $t$  in a directed (resp. undirected) graph. These two graph connectivity problems played a central role in understanding the complexity classes  $\mathbf{L}$ ,  $\mathbf{SL}$  and  $\mathbf{NL}$  [LP82, NSW92, SZ99, ATWZ00, NT95, AKL<sup>+</sup>79, BCD<sup>+</sup>89, KW93, Tri08, RVW00, Rei08].

In Section 2, we introduce a new graph connectivity problem, which we call  $\mathbf{ST-REALIZABILITY}$  and prove that  $\mathbf{ST-REALIZABILITY}$  is complete for  $\mathbf{LogCFL}$ .  $\mathbf{ST-REALIZABILITY}$  is a generalization of  $\mathbf{ST-CONNECTIVITY}$ , which is  $\mathbf{NL}$ -complete. Our definition of  $\mathbf{ST-REALIZABILITY}$  is motivated by (i) Hardest CFL [Gre73, Sud78, Har78], (ii) Labeled Acyclic GAP, which is  $\mathbf{LogCFL}$ -complete [GHR95] (iii) CFL-reachability, which is  $\mathbf{P}$ -complete [MR00, AP87, Rep96, UG86] and (iv) the insights from Niedermeier and Rossmanith's parsimonious simulation of  $\mathbf{LogCFL}$  by  $\mathbf{SAC}^1$  circuits [NR95].

Unlike  $\mathbf{ST-CONNECTIVITY}$ , using breadth-first search (or) depth-first search and keeping track of “visited” vertices does not result in a polynomial time algorithm for  $\mathbf{ST-REALIZABILITY}$ . In Section 5, we generalize the notions of transitive closure and graph squaring. Using these generalizations we present a natural polynomial time algorithm to compute the generalized transitive closure, thus solving  $\mathbf{ST-REALIZABILITY}$ .

**Symmetric AuxPDAs** : In Section 3, we define  $\mathbf{UNDIRECTED ST-REALIZABILITY}$ , a “symmetric” version of  $\mathbf{ST-REALIZABILITY}$ . To study the space complexity of  $\mathbf{UNDIRECTED ST-REALIZABILITY}$  we define *symmetric* auxiliary pushdown automata, a natural generalization of symmetric Turing machines, introduced by Lewis and Papadimitriou [LP82]. We introduce a new complexity class called  $\mathbf{SLogCFL}$ , a generaliza-

tion of **SL** and show that  $\mathbf{LogDCFL} \subseteq \mathbf{SLogCFL} \subseteq \mathbf{LogCFL}$ .

**Graph Realizability Problems** : In Section 4, we study several variants of **ST-REALIZABILITY** and the corresponding complexity classes. All of these complexity classes lie between **L** and **LogCFL**. In particular, **BALANCED ST-CONNECTIVITY** and **POSITIVE BALANCED ST-CONNECTIVITY** are natural graph connectivity problems that lie between **L** and **NL**. Figure 1 summarizes the relationship among the newly defined classes.

**Space Efficient Algorithms** : The **L** vs **SL** question (i.e., is there a log space algorithm for solving **UNDIRECTED ST-CONNECTIVITY**) motivated an exciting series of new concepts and techniques. Prior to the work of Lewis and Papadimitriou [LP82], Aleliunas et. al. [AKL<sup>+</sup>79] proved that **UNDIRECTED ST-CONNECTIVITY**  $\in$  **RL**, implying  $\mathbf{SL} \subseteq \mathbf{RL}$ . Nisan, Szemerédi and Wigderson [NSW92] showed that **UNDIRECTED ST-CONNECTIVITY** can be solved deterministically in space  $O(\log^{\frac{3}{2}} n)$ . This result was later subsumed by a beautiful result of Saks and Zhou, showing that  $BP_HSPACE(S) \subseteq DSPACE(S^{3/2})$  [SZ99]. Armoni, et. al. [ATWZ00] showed that **UNDIRECTED ST-CONNECTIVITY**  $\in$   $DSPACE(\log^{\frac{4}{3}} n)$ . Trifonov [Tri08] gave an  $O(\log n \log \log n)$ -space deterministic algorithm for **UNDIRECTED ST-CONNECTIVITY**. Independently at the same time, using completely different techniques, Reingold [Rei08] settled the space complexity of **UNDIRECTED ST-CONNECTIVITY** and proved that  $\mathbf{SL} = \mathbf{L}$ . The zig-zag graph product, introduced by Reingold, Vadhan and Wigderson [RVW02], played a crucial role in Reingold’s algorithm.

Our space efficient algorithm for **SGSLogCFL** (see Section 7) is based on Trifonov’s technique [Tri08], which is based on Chong-Lam’s parallel algorithm [CL95] solving **UNDIRECTED ST-CONNECTIVITY** in  $O(\log n \log \log n)$  time on EREW PRAM. This necessitates the development of such a parallel algorithm for **SGSLogCFL**.

**Parallel Algorithms** : Hirschberg, Chandra and Sarwate [HCS79] presented an  $O(\log^2 n)$  time parallel algorithm using  $n^2/\log n$  processors on a CREW PRAM to find connected components of an undirected graph. Their algorithm remained the best known for almost a decade. In a breakthrough work, Johnson and Metaxas [JM97] presented a CREW algorithm running in  $O(\log^{\frac{3}{2}} n)$  time using  $n + m$  processors. Subsequently they improved their algorithm to run on an EREW PRAM with the same time complexity and number of processors [JM95]. Chong and Lam [CL95] presented an  $O(\log n \log \log n)$  time deterministic EREW PRAM algorithm with  $O(m + n)$  processors. Chong, Han, and Lam [CHL99] showed that the problem can be solved on the EREW PRAM in  $O(\log n)$  time with  $O(m + n)$  processors.

In Section 6, we generalize the algorithms of [HCS79], [JM97] and [CL95] and design the corresponding parallel algorithms for **SGSLogCFL**. In Section 7, we use these algorithms to prove that **SGSLogCFL** is contained in  $DSPACE(\log n \log \log n)$ .

## 2 Realizable Paths

### 2.1 ST-REALIZABILITY

We are given a directed graph  $\mathcal{G}(V, E)$ , a vertex labeling function  $L_V : V \rightarrow \{\alpha_1, \alpha_2, \dots, \alpha_k\}$  and an edge labeling function  $L_E : E \rightarrow \{\text{push}, \text{pop}, \epsilon\}$ . The ordered pair  $(s, t)$ , where  $s, t \in V$ , is said to be **realizable** if the following two conditions hold :

- There is a directed path (say  $P$ ) from  $s$  to  $t$ .

- The concatenation of the vertex and edge labels along the path  $P$  is a *realizable* string (see Definition 2.1).

**Definition 2.1.** Let  $\mathcal{A} = \{push, pop, \epsilon, \alpha_1, \alpha_2, \dots, \alpha_k\}$  be the set of alphabets. A **realizable string** is a nonempty string of alphabets from  $\mathcal{A}$ , defined in the following recursive manner :

- for all  $1 \leq i \leq k$ , “ $\alpha_i$ ” is a realizable string.
- for all  $1 \leq i \leq k$ , “ $\alpha_i \epsilon \alpha_i$ ” is a realizable string.
- if  $S$  is a realizable string then so is “ $\alpha_i push S pop \alpha_i$ ”, for all  $1 \leq i \leq k$ .
- for all  $1 \leq i \leq k$ , if “ $\alpha_i S_1 \alpha_i$ ” and “ $\alpha_i S_2 \alpha_i$ ” are realizable strings then so is “ $\alpha_i S_1 \alpha_i S_2 \alpha_i$ ”.

ST-REALIZABILITY : Given a directed graph  $\mathcal{G}(V, E)$  with vertices labeled from  $\{\alpha_1, \alpha_2, \dots, \alpha_k\}$  and edges labeled from  $\{push, pop, \epsilon\}$  and two distinguished nodes  $s$  and  $t$ , decide if there is a realizable path from  $s$  to  $t$  in  $\mathcal{G}$ .

We use the notation  $(u \rightsquigarrow v)$  to denote that there is a realizable path from  $u$  to  $v$ . If all the vertices of  $\mathcal{G}$  are labeled  $\alpha_1$  (i.e.,  $k = 1$ ) and all the edges are labeled  $\epsilon$ , we get an instance of ST-CONNECTIVITY. Hence, ST-REALIZABILITY is a generalization of ST-CONNECTIVITY.

**Theorem 2.2.** ST-REALIZABILITY is **LogCFL**-complete.

**Corollary 2.3.** ST-REALIZABILITY with no  $\epsilon$ -edges is **LogCFL**-complete.

## 2.2 Graph Representation

We now discuss the representation of an instance of ST-REALIZABILITY i.e., a directed graph  $\mathcal{G}$  with the vertex and edge labels. Let this graph be  $\mathcal{G}(V, E)$  with  $|V| = n$ . For simplicity we assume that there are no multi-edges. We represent  $\mathcal{G}$  as a 4-tuple  $\mathcal{G} = \langle \mathcal{L}, \mathcal{P}_{push}, \mathcal{P}_{pop}, \mathcal{E} \rangle$ , where  $\mathcal{L}$  is an integer array of length  $n$ ,  $\mathcal{P}_{push}$ ,  $\mathcal{P}_{pop}$  and  $\mathcal{E}$  are  $n \times n$  boolean matrices.  $\mathcal{L}$  is an integer array of length  $N$  representing the vertex labels.  $\mathcal{L}[u]$  represents the label of vertex  $u$  i.e.,  $\mathcal{L}[u] = i$  iff the label of  $u$  is  $\alpha_i$ . The  $[u, v]^{th}$  entry of the matrix  $\mathcal{P}_{push}$  (resp.  $\mathcal{P}_{pop}$  and  $\mathcal{E}$ ) is 1 if and only if the directed edge  $(u, v)$  is labeled *push* (resp. *pop* and  $\epsilon$ ). We may assume that  $L_E(u, u) = \epsilon$  for all  $u \in V$  i.e.,  $\mathcal{E}[u, u] = \epsilon$  for all  $u \in V$ .

## 2.3 Gap Matrix

**Definition 2.4.** (Niedermeier and Rossmanith [NR95]) : Let  $a, b, c, d$  be four configurations such that :  $a$  and  $b$  have same pushdown heights,  $c$  and  $d$  have same pushdown heights and there exists a computation path from  $a$  to  $c$  and one from  $d$  to  $b$ . The level of the pushdown must not go below the level of  $a$  and  $b$  during the computation. We say that  $(a, b)$  is *realizable with gap*  $(c, d)$ .

In the context of ST-REALIZABILITY, we relax the above definition as shown below. This allows us to define a natural repeated squaring algorithm to solve ST-REALIZABILITY. For the rest of this paper, we will use the following definition.

**Path with gap** : A *path with gap* consists of four vertices  $a, b, c, d$  such that (i) there is a computation path  $P_1$  from  $a$  to  $c$  and  $P_2$  from  $d$  to  $b$  (ii) the vertex labels of  $a$  and  $b$  are the same (iii) the vertex labels of  $c$  and  $d$  are the same (iv) let  $P$  be the path formed by concatenating  $P_1$  and  $P_2$  i.e., identifying  $c$  and  $d$  (v) the concatenation of the vertex and edge labels along the path  $P$  is a *realizable* string. We denote such a “path with gap” by  $(a \rightsquigarrow (c, d) \rightsquigarrow b)$  and say that  $(a, b)$  is *realizable with gap*  $(c, d)$ .

*Pair-with-gap*  $(a \rightsquigarrow (c, d) \rightsquigarrow b)$  is interpreted as if the two surface configurations  $c$  and  $d$  were the same, i.e., as if a realizable path from  $c$  to  $d$  would exist. To keep track of paths with gaps, we maintain a boolean *gap matrix*  $\Upsilon$ , indexed by 4-tuple of vertices  $[a, (c, d), b]$  such that if  $\Upsilon[a, (c, d), b] = 1$  then  $(a \rightsquigarrow (c, d) \rightsquigarrow b)$ . We initialize the gap matrix  $\Upsilon$  with the labels from the matrices  $\mathcal{L}, \mathcal{P}_{push}$  and  $\mathcal{P}_{pop}$  as follows.

---

**InitializeGapMatrix**( $\Upsilon$ )

for all  $a, b, c, d \in V$   $\Upsilon[a, (c, d), b] = 0$

for all  $a, b, c, d \in V$

**if**  $((\mathcal{P}_{push}[a, c] == 1) \&\& (\mathcal{P}_{pop}[d, b] == 1) \&\& (\mathcal{L}[a] == \mathcal{L}[b]) \&\& (\mathcal{L}[c] == \mathcal{L}[d]))$

**then**  $\Upsilon[a, (c, d), b] = 1$

for all  $a \in V$   $\Upsilon[a, (a, a), a] = 1$

for all  $a, b \in V$   $\Upsilon[a, (a, b), b] = 1$

---

All the required information from the matrices  $\mathcal{L}, \mathcal{P}_{push}$  and  $\mathcal{P}_{pop}$  is now present in the gap matrix  $\Upsilon$ . Note that we are implicitly removing the “unnecessary” edges as follows.

**Removing unnecessary edges** : If  $s$  and  $t$  are realizable in  $\mathcal{G}$  along a path  $P$  then the *push* and *pop* edges along  $P$  have to “match” i.e., every *push* label has a corresponding *pop* label. In other words, if there is a *push* edge  $(a, c)$  such that the label of  $a$  is  $\alpha_i$  and the label of  $c$  is  $\alpha_j$  then there is a corresponding *pop* edge  $(d, b)$  along the path  $P$  such that the label of  $d$  is  $\alpha_j$  and the label of  $b$  is  $\alpha_i$ . Hence, we can remove the unnecessary edges as follows :

- Let  $(u, v)$  be a *push* edge in  $\mathcal{G}$  such that the label of  $u$  is  $\alpha_i$  and the label of  $v$  is  $\alpha_j$ . If there is no *pop* edge in  $\mathcal{G}$  (other than  $(v, u)$ ) with the vertex labels  $(\alpha_j, \alpha_i)$ , then remove the edge  $(u, v)$ .
- Let  $(u, v)$  be a *pop* edge in  $\mathcal{G}$  such that the label of  $u$  is  $\alpha_i$  and the label of  $v$  is  $\alpha_j$ . If there is no *push* edge in  $\mathcal{G}$  (other than  $(v, u)$ ) with the vertex labels  $(\alpha_j, \alpha_i)$ , then remove the edge  $(u, v)$ .

We call  $\mathcal{E}$  the *standard* matrix and  $\Upsilon$  the *gap* matrix and assume that an instance of ST-REALIZABILITY,  $\mathcal{H}$ , is represented by an  $n \times n$  standard matrix  $\mathcal{E}$  and an  $n^2 \times n^2$  gap matrix  $\Upsilon$  and denote this by  $\mathcal{H} = \langle \Upsilon, \mathcal{E} \rangle$ . The rows and columns of  $\Upsilon$  are indexed by pairs of vertices of  $\mathcal{H}$ .  $\Upsilon[a, (c, d), b]$  corresponds to the  $[(a, b), (c, d)]^{th}$  entry in the  $n^2 \times n^2$  matrix.

### 3 UNDIRECTED ST-REALIZABILITY and Symmetric AuxPDAs

#### 3.1 UNDIRECTED ST-REALIZABILITY

We are given an undirected graph  $\mathcal{G}(V, E)$ , a vertex labeling function  $L_V : V \rightarrow \{\alpha_1, \alpha_2, \dots, \alpha_k\}$  and an edge labeling function  $L_E : E \rightarrow \{push, pop, \epsilon\}$ . Moreover, the edge labels are “symmetric” i.e., they satisfy

the following properties : (i)  $L_E(u, v) = push$  if and only if  $L_E(v, u) = pop$  and (ii)  $L_E(u, v) = \epsilon$  if and only if  $L_E(v, u) = \epsilon$ .

The pair  $(s, t)$ , where  $s, t \in V$ , is said to be *realizable* if there is an undirected path (say  $P$ ) from  $s$  to  $t$  and the concatenation of the vertex and edge labels along the path  $P$  is a *realizable* string. Since the edge labels are symmetric,  $(s, t)$  is realizable if and only if  $(t, s)$  is realizable. We denote this by  $(s \leftrightarrow t)$ .

**UNDIRECTED ST-REALIZABILITY** : Given an undirected graph  $\mathcal{G}(V, E)$  with vertices labeled from  $\{\alpha_1, \alpha_2, \dots, \alpha_k\}$  and *symmetric* edge labels from  $\{push, pop, \epsilon\}$  and two distinguished nodes  $s$  and  $t$ , decide if  $s$  and  $t$  are realizable in  $\mathcal{G}$ .

If all the vertices of  $\mathcal{G}$  are labeled  $\alpha_1$  (i.e.,  $k = 1$ ) and all the edges are labeled  $\epsilon$ , we get an instance of **UNDIRECTED ST-CONNECTIVITY**. Hence, **UNDIRECTED ST-REALIZABILITY** is a generalization of **UNDIRECTED ST-CONNECTIVITY**. To study the space complexity of **UNDIRECTED ST-REALIZABILITY** we introduce *symmetric* AuxPDAs in the following subsection.

### 3.2 Symmetric AuxPDAs

Intuitively, a *symmetric* AuxPDA is a nondeterministic multi-tape Turing machine which has an extra tape called pushdown tape, with an additional requirement that every move of the machine is “reversible”. In other words, the “yields” relation between its (surface) configurations is symmetric. Such a machine is allowed to scan two symbols at a time on each of its tapes. We present the formal definitions, properties of symmetric AuxPDAs and the proofs of the following theorems in the **appendix** (see Appendix A). We define **SLogCFL** to be the class of languages accepted by a log space bounded and polynomial time bounded symmetric AuxPDA.

**Theorem 3.1.**  $\text{LogDCFL} \subseteq \text{SLogCFL} \subseteq \text{LogCFL}$ .

**Theorem 3.2.** **UNDIRECTED ST-REALIZABILITY** is **SLogCFL**-complete.

**Corollary 3.3.** **UNDIRECTED ST-REALIZABILITY** with no  $\epsilon$ -edges is **SLogCFL**-complete.

## 4 More Realizability Problems between L and LogCFL

As noted earlier, an instance  $\mathcal{H}$  of **ST-REALIZABILITY** is represented by an  $n \times n$  standard matrix  $\mathcal{E}$  and an  $n^2 \times n^2$  gap matrix  $\Upsilon$ . The vertices of  $\mathcal{H}$  are labeled with  $\{\alpha_1, \dots, \alpha_k\}$ . In this section, we define more graph realizability problems based on the symmetry of the matrices  $\Upsilon$  and  $\mathcal{E}$  and the number of distinct vertex labels (i.e., number of stack symbols, denoted by  $k$ ). We define the corresponding complexity classes as the set of all languages that are logspace reducible to the corresponding graph realizability problem. Table 1 summarizes all the definitions. The prefix **S** is used to denote the symmetry of the standard matrix. The prefix **SGS** is used to denote the symmetry of the standard *and* gap matrix. A moment of thought would reveal that the case of symmetric gap matrix and asymmetric standard matrix does not make much sense. The prefix **1** is used to denote that there is only one vertex label.

Complexity class	Number of stack symbols	Standard Matrix	Gap Matrix
<b>LogCFL</b>	$k \geq 2$	asymmetric	asymmetric
<b>SLogCFL</b>	$k \geq 2$	symmetric	asymmetric
<b>SGSLogCFL</b>	$k \geq 2$	symmetric	symmetric
<b>1LogCFL</b>	$k = 1$	asymmetric	asymmetric
<b>1SLogCFL</b>	$k = 1$	symmetric	asymmetric
<b>1SGSLogCFL</b>	$k = 1$	symmetric	symmetric

Table 1: Graph realizability problems between **L** and **LogCFL**.

#### 4.1 Realizability with Symmetric Gap

We are given an undirected graph  $\mathcal{G}(V, E)$ , a vertex labeling function  $L_V : V \rightarrow \{\alpha_1, \alpha_2, \dots, \alpha_k\}$  and an edge labeling function  $L_E : E \rightarrow \{push, pop, \epsilon\}$ . The edge labels are “symmetric” as defined in Section 3. The pair  $(s, t)$ , where  $s, t \in V$ , is said to be **realizable with symmetric gap** if the following two conditions hold :

- There is an undirected path (say  $P$ ) from  $s$  to  $t$ .
- The concatenation of the vertex and edge labels along the path  $P$  is a *realizable string with symmetric gap* (see Definition 4.1).

**Definition 4.1.** Let  $\mathcal{A} = \{push, pop, \epsilon, \alpha_1, \alpha_2, \dots, \alpha_k\}$  be the set of alphabets. A **realizable string with symmetric gap** is a nonempty string of alphabets from  $\mathcal{A}$ , defined in the following recursive manner :

- for all  $1 \leq i \leq k$ , “ $\alpha_i$ ” is a realizable string.
- for all  $1 \leq i \leq k$ , “ $\alpha_i \epsilon \alpha_i$ ” is a realizable string.
- if  $S$  is a realizable string then so is “ $\alpha_i push S pop \alpha_i$ ”, for all  $1 \leq i \leq k$ .
- if  $S$  is a realizable string then so is “ $\alpha_i pop S push \alpha_i$ ”, for all  $1 \leq i \leq k$ .
- for all  $1 \leq i \leq k$ , if “ $\alpha_i S_1 \alpha_i$ ” and “ $\alpha_i S_2 \alpha_i$ ” are realizable strings then so is “ $\alpha_i S_1 \alpha_i S_2 \alpha_i$ ”.

Since the edge labels are symmetric,  $(s, t)$  is realizable if and only if  $(t, s)$  is realizable. We initialize the gap matrix as described in Section 2.3. By the definition of *realizable string with symmetric gap*,  $(a \rightsquigarrow (c, d) \rightsquigarrow b)$  if and only if  $(c \rightsquigarrow (a, b) \rightsquigarrow d)$ . Hence the corresponding  $n^2 \times n^2$  gap matrix  $\Upsilon$  is a symmetric matrix. We denote this symmetry by  $(a \rightsquigarrow (c, d) \rightsquigarrow b)$ .

**SYMMETRIC GAP UNDIRECTED ST-REALIZABILITY :** Given an undirected graph  $\mathcal{G}(V, E)$  with vertices labeled from  $\{\alpha_1, \alpha_2, \dots, \alpha_k\}$  and *symmetric* edge labels from  $\{push, pop, \epsilon\}$  and two distinguished nodes  $s$  and  $t$ , decide if  $s$  and  $t$  are *realizable with symmetric gap* in  $\mathcal{G}$ .

**SGSLogCFL** is the class of languages that are logspace reducible to SYMMETRIC GAP UNDIRECTED ST-REALIZABILITY.



## 4.2 Realizability with one stack symbol

The complexity classes  $1\text{LogCFL}$ ,  $1\text{SLogCFL}$  and  $1\text{SGSLogCFL}$  are obtained by restricting  $\text{LogCFL}$ ,  $\text{SLogCFL}$  and  $\text{SGSLogCFL}$  respectively to use only one stack symbol i.e., by insisting that  $k = 1$  in the above definitions. Since the vertices are all labeled with one label, we may omit the vertex labels in the definitions. After omitting the vertex labels, the corresponding *realizability* can be defined using a context-free language as shown below.

### 4.2.1 $1\text{LogCFL}$

$1\text{LogCFL}$  is the class of languages that are logspace reducible to the following graph realizability problem. We are given a directed graph  $\mathcal{G}(V, E)$ , with edges labeled from  $\{\text{push}, \text{pop}, \epsilon\}$ . The ordered pair  $(s, t)$ , where  $s, t \in V$ , is said to be *realizable* if the following two conditions hold :

- There is a directed path (say  $P$ ) from  $s$  to  $t$ .
- The concatenation of the edge labels on the path  $P$  is a string produced by the following context-free grammar :  $S \rightarrow S S; S \rightarrow \text{push } S \text{ pop}; S \rightarrow \epsilon; S \rightarrow \emptyset$ . Here  $\emptyset$  denotes the empty string.

### 4.2.2 $1\text{SLogCFL}$

We are given an undirected graph  $\mathcal{G}(V, E)$ , with the edges labeled from  $\{\text{push}, \text{pop}, \epsilon\}$ . Moreover, the edge labels are “symmetric” as defined in Section 3. The pair  $(s, t)$ , where  $s, t \in V$ , is said to be *realizable* if there is an undirected path (say  $P$ ) from  $s$  to  $t$  and the concatenation of the edge labels along the path  $P$  is a string produced by the context-free grammar mentioned in Section 4.2.1. Since the edge labels are symmetric,  $(s, t)$  is realizable if and only if  $(t, s)$  is realizable.  $1\text{SLogCFL}$  is the class of languages that are logspace reducible to this undirected graph realizability problem.

### 4.2.3 $1\text{SGSLogCFL}$

$1\text{SGSLogCFL}$  is the class of languages that are logspace reducible to the following graph realizability problem. We are given an undirected graph  $\mathcal{G}(V, E)$ , with the edges labeled from  $\{\text{push}, \text{pop}, \epsilon\}$ . The edge labels are “symmetric” as defined in Section 3. The pair  $(s, t)$ , where  $s, t \in V$ , is said to be *realizable* if the following two conditions hold :

- There is a simple undirected path (say  $P$ ) from  $s$  to  $t$ .
- The concatenation of the edge labels on the path  $P$  is a string produced by the following context-free grammar :  $S \rightarrow S S; S \rightarrow \text{push } S \text{ pop}; S \rightarrow \text{pop } S \text{ push}; S \rightarrow \epsilon; S \rightarrow \emptyset$ . Here  $\emptyset$  denotes the empty string.

## 4.3 Relationship among the Realizability Problems

By definition, we have the following inclusions.

- $\text{SGSLogCFL} \subseteq \text{SLogCFL} \subseteq \text{LogCFL}$
- $1\text{SGSLogCFL} \subseteq 1\text{SLogCFL} \subseteq 1\text{LogCFL}$
- $1\text{LogCFL} \subseteq \text{LogCFL}$

- $1\text{SLogCFL} \subseteq \text{SLogCFL}$
- $1\text{SGSLogCFL} \subseteq \text{SGSLogCFL}$

Independent to our work, Allender and Lange [AL10] defined symmetric AuxPDAs and proved that every language accepted by a nondeterministic auxiliary pushdown automaton in polynomial time can be accepted by a symmetric auxiliary pushdown automaton in polynomial time. Their definition of symmetric AuxPDAs is equivalent to ours [All]. Borodin et. al. [BCD<sup>+</sup>89] proved that  $\text{LogCFL} = \text{co-LogCFL}$ . The following theorem and its corollary are immediate.

**Theorem 4.2.** (Allender and Lange [AL10]).  $\text{SLogCFL} = \text{LogCFL}$ .

**Corollary 4.3.**  $\text{SLogCFL} = \text{co-SLogCFL}$ .

#### 4.4 Realizability Problems between L and NL

All the realizability problems defined above are generalizations of  $\text{UNDIRECTED ST-CONNECTIVITY}$ . Hence, the corresponding complexity classes contain  $\text{L}$ . We now prove that  $\text{NL} = 1\text{LogCFL}$ . Hence,  $\text{L} = \text{SL} \subseteq 1\text{SGSLogCFL} \subseteq 1\text{SLogCFL} \subseteq 1\text{LogCFL} = \text{NL}$ . We introduce two natural graph connectivity problems characterizing  $1\text{SGSLogCFL}$  and  $1\text{SLogCFL}$ .

**Theorem 4.4.**  $\text{NL} = 1\text{LogCFL}$ .

**Corollary 4.5.**  $\text{L} = \text{SL} \subseteq 1\text{SGSLogCFL} \subseteq 1\text{SLogCFL} \subseteq 1\text{LogCFL} = \text{NL}$ .

Let  $\mathcal{G}(V, E)$  be a directed graph. Let  $\mathcal{G}'(V, E')$  be the underlying undirected graph of  $\mathcal{G}$ . Let  $P$  be a path in  $\mathcal{G}'$ . Let  $e = (u, v)$  be an edge along the path  $P$ . Edge  $e$  is called *neutral* edge if both  $(u, v)$  and  $(v, u)$  are in  $E$ . Edge  $e$  is called *forward* edge if  $(u, v) \in E$  and  $(v, u) \notin E$ . Edge  $e$  is called *backward* edge if  $(u, v) \notin E$  and  $(v, u) \in E$ .

A path (say  $P$ ) from  $s \in V$  to  $t \in V$  in  $\mathcal{G}'(V, E')$  is called *balanced* if the number of forward edges along  $P$  is equal to the number of backward edges along  $P$ . A balanced path might have any number of neutral edges. By definition, if there is a balanced path from  $s$  to  $t$  then there is a balanced path from  $t$  to  $s$ . The path  $P$  may not be a simple path. We are concerned with balanced paths of length at most  $n$ . See Section C in the **appendix** for more details and variants of balanced connectivity problems.

**BALANCED ST-CONNECTIVITY :** Given a directed graph  $\mathcal{G}(V, E)$  and two distinguished nodes  $s$  and  $t$ , decide if there is *balanced* path (of length at most  $n$ ) between  $s$  and  $t$ .

Let  $P$  be a path from  $s \in V$  to  $t \in V$  in  $\mathcal{G}(V, E)$ . We say  $v \in P$  if the vertex  $v$  is on the path  $P$ . For  $v \in P$  we denote by  $P_v$  the subpath of  $P$  starting from  $s$  and ending at  $v$ . We say that  $P$  is *positive* if the number of forward edges of  $P_v$  is at least the number of backward edges of  $P_v$ , for all  $v \in P$ . In other words, the number of forward edges minus the number of backward edges of  $P_v$  is positive, for all  $v \in P$ . We say that  $P$  is *positive balanced* if  $P$  is positive and balanced. By definition, if there is a positive balanced path from  $s$  to  $t$  then there is a positive balanced path from  $t$  to  $s$ .

POSITIVE BALANCED ST-CONNECTIVITY : Given a directed graph  $\mathcal{G}(V, E)$  and two distinguished nodes  $s$  and  $t$ , decide if there is *positive balanced* path (of length at most  $n$ ) between  $s$  and  $t$ .

**Theorem 4.6.** BALANCED ST-CONNECTIVITY is **1SGSLogCFL**-complete.

**Theorem 4.7.** POSITIVE BALANCED ST-CONNECTIVITY is **1SLogCFL**-complete.

Figure 1 summarizes the relationship among the above defined classes.

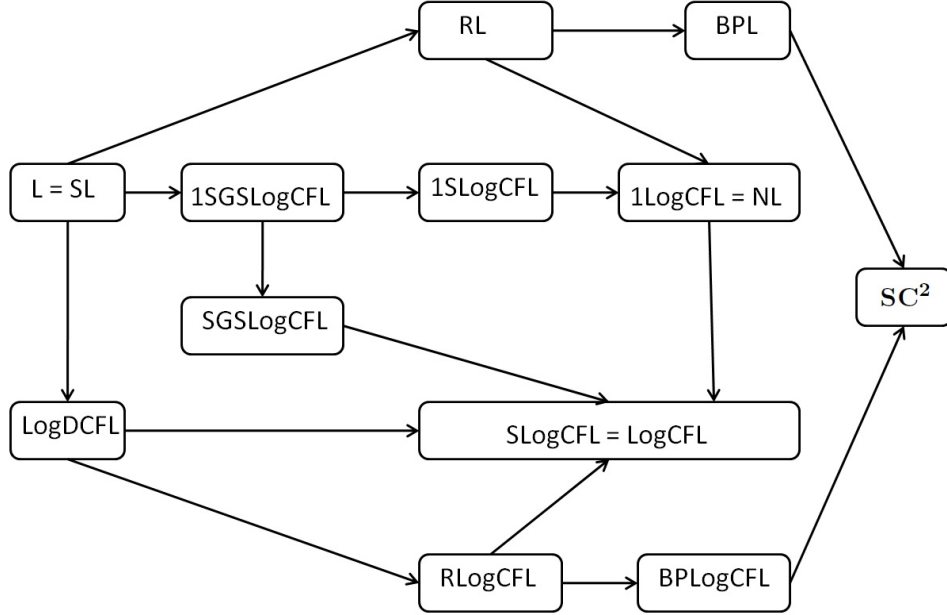


Figure 1: Relationship between complexity classes. A directed edge from class **A** to class **B** shows that  $\mathbf{A} \subseteq \mathbf{B}$ . In addition,  $\mathbf{RL} \subseteq \mathbf{RLogCFL}$  and  $\mathbf{BPL} \subseteq \mathbf{BPLogCFL}$ . BALANCED ST-CONNECTIVITY is **1SGSLogCFL**-complete and POSITIVE BALANCED ST-CONNECTIVITY is **1SLogCFL**-complete.

## 5 Transitive Closure

The definitions and theorems in this section apply to all the graph realizability problems defined above. We present the definitions and theorems for ST-REALIZABILITY, the most general graph realizability problem.

**Definition 5.1.** Let  $\mathcal{G} = \langle \Upsilon, \mathcal{E} \rangle$  be an instance of ST-REALIZABILITY. The **transitive closure** of  $\mathcal{G}$ , denoted by  $\mathcal{G}^* = \langle \Upsilon^*, \mathcal{E}^* \rangle$ , is a pair of gap and standard matrix such that for all  $a, b, c, d \in V$ ,

- (i)  $\mathcal{E}^*[a][b] = 1$  iff  $(a \rightsquigarrow b)$  and
- (ii)  $\Upsilon^*[a, (c, d), b] = 1$  iff  $(a, b)$  is realizable with gap  $(c, d)$ .

## 5.1 Tensor Products

We now present several tensor products acting on  $\mathcal{E}$  and  $\Upsilon$ . The products  $\otimes_1$  to  $\otimes_5$  are introduced in [Ven06]. We introduce  $\otimes_6$  and  $\otimes_7$ . These products update the standard matrix  $\mathcal{E}$  and the gap matrix  $\Upsilon$  with new “connectivity information” of  $\mathcal{G}$ . Let  $\mathcal{E}, \mathcal{E}_1, \mathcal{E}_2$  represent standard matrices and  $\Upsilon, \Upsilon_1, \Upsilon_2$  represent gap matrices. Let  $a, b, c, d, z$  represent the vertices of  $\mathcal{G}$ . Matrices indexed by two (resp. four) indices are standard (resp. gap) matrices. When we are dealing with boolean matrices, all the summations (resp. multiplications) are interpreted as boolean  $\vee$  (resp. boolean  $\wedge$ ).

1. If  $(a \rightsquigarrow z)$  and  $(z \rightsquigarrow b)$  then  $(a \rightsquigarrow b)$  :

$$(\mathcal{E}_1 \otimes_1 \mathcal{E}_2)[a, b] = \sum_z \mathcal{E}_1[a, z] \cdot \mathcal{E}_2[z, b].$$

2. If  $(a \rightsquigarrow (c, d) \rightsquigarrow b)$  and  $(c \rightsquigarrow d)$  then  $(a \rightsquigarrow b)$  :

$$(\Upsilon \otimes_2 \mathcal{E})[a, b] = \sum_{c, d} \Upsilon[a, (c, d), b] \cdot \mathcal{E}[c, d].$$

3. If  $(a \rightsquigarrow (c, d) \rightsquigarrow b)$  and  $(b \rightsquigarrow z)$  then  $(a \rightsquigarrow (c, d) \rightsquigarrow z)$  :

$$(\Upsilon \otimes_3 \mathcal{E})[a, (c, d), z] = \sum_b \Upsilon[a, (c, d), b] \cdot \mathcal{E}[b, z].$$

4. If  $(z \rightsquigarrow a)$  and  $(a \rightsquigarrow (c, d) \rightsquigarrow b)$  then  $(z \rightsquigarrow (c, d) \rightsquigarrow b)$  :

$$(\mathcal{E} \otimes_4 \Upsilon)[z, (c, d), b] = \sum_a \mathcal{E}[z, a] \cdot \Upsilon[a, (c, d), b].$$

5. If  $(a \rightsquigarrow (c, d) \rightsquigarrow b)$  and  $(c \rightsquigarrow (e, f) \rightsquigarrow d)$  then  $(a \rightsquigarrow (e, f) \rightsquigarrow b)$  :

$$(\Upsilon_1 \otimes_5 \Upsilon_2)[a, (e, f), b] = \sum_{c, d} \Upsilon_1[a, (c, d), b] \cdot \Upsilon_2[c, (e, f), d].$$

6. If  $(a \rightsquigarrow (c, d) \rightsquigarrow b)$  and  $(z \rightsquigarrow d)$  then  $(a \rightsquigarrow (c, z) \rightsquigarrow b)$  :

$$(\Upsilon \otimes_6 \mathcal{E})[a, (c, z), b] = \sum_d \Upsilon[a, (c, d), b] \cdot \mathcal{E}[z, d].$$

7. If  $(a \rightsquigarrow (c, d) \rightsquigarrow b)$  and  $(c \rightsquigarrow z)$  then  $(a \rightsquigarrow (z, d) \rightsquigarrow b)$  :

$$(\Upsilon \otimes_7 \mathcal{E})[a, (z, d), b] = \sum_c \Upsilon[a, (c, d), b] \cdot \mathcal{E}[c, z].$$

## 5.2 Computing Transitive Closure

Given  $\mathcal{G} = \langle \Upsilon, \mathcal{E} \rangle$  the following algorithm computes  $\mathbf{Square}(\mathcal{G})$ . This algorithm is based on a parsimonious simulation of **LogCFL** by  $\mathbf{SAC}^1$  circuits given by Niedermeier and Rossmanith [NR95]. Implementation of  $\mathbf{Square}(\langle \Upsilon, \mathcal{E} \rangle)$  using the above mentioned tensor products is shown below. The following theorem implies a natural polynomial time algorithm to solve **ST-REALIZABILITY**.

---

**Square**( $\langle \Upsilon, \mathcal{E} \rangle$ )

for all  $a, b \in V$  update  $\mathcal{E}$  as follows :

$$\mathcal{E}[a, b] = \sum_{c, e, f, g, d} \Upsilon[a, (c, d), b] \cdot \Upsilon[c, (e, f), g] \cdot \mathcal{E}[e, f] \cdot \mathcal{E}[g, d]$$

for all  $a, b, c, d \in V$  update  $\Upsilon$  as follows :

$$\begin{aligned} \Upsilon[a, (c, d), b] &= \sum_{c', e', f', g', d'} \Upsilon[a, (c'd'), b] \cdot \Upsilon[c', (e', f'), g'] \cdot \Upsilon[e', (c, d), f'] \cdot \mathcal{E}[g', d'] \\ &+ \sum_{c', e', f', g', d'} \Upsilon[a, (c'd'), b] \cdot \Upsilon[c', (e', f'), g'] \cdot \mathcal{E}[e', f'] \cdot \Upsilon[g', (c, d), d'] \end{aligned}$$

**return**  $\langle \Upsilon, \mathcal{E} \rangle$

---



---

**Square**( $\langle \Upsilon, \mathcal{E} \rangle$ )

$$\mathcal{E} = (\Upsilon \otimes_2 ((\Upsilon \otimes_2 \mathcal{E}) \otimes_1 \mathcal{E}))$$

$$\Upsilon = (\Upsilon \otimes_5 ((\Upsilon \otimes_5 \Upsilon) \otimes_3 \mathcal{E})) + (\Upsilon \otimes_5 ((\Upsilon \otimes_2 \mathcal{E}) \otimes_4 \Upsilon))$$

**return**  $\langle \Upsilon, \mathcal{E} \rangle$

---

**Theorem 5.2.** Let  $\mathcal{G}$  be an instance of **ST-REALIZABILITY**.  $\mathcal{G}^* = \langle \Upsilon^*, \mathcal{E}^* \rangle$  can be computed using  $O(\log n)$  repeated applications of  $\mathbf{Square}(\mathcal{G})$ .

## 5.3 Simple Squaring Operation

The following algorithm **SimpleSquare** is a more intuitive squaring operation. It plays a crucial role in the proofs of correctness of parallel and space efficient algorithms for **SGSLogCFL** (see Section 6 and Section 7).

---

**SimpleSquare**( $\langle \Upsilon, \mathcal{E} \rangle$ )

$$\mathcal{E} = \mathcal{E} \otimes_1 \mathcal{E}$$

$$\mathcal{E} = \Upsilon \otimes_2 \mathcal{E}$$

$$\Upsilon = \Upsilon \otimes_3 \mathcal{E}$$

$$\Upsilon = \mathcal{E} \otimes_4 \Upsilon$$

$$\Upsilon = \Upsilon \otimes_5 \Upsilon$$

$$\Upsilon = \Upsilon \otimes_6 \mathcal{E}$$

$$\Upsilon = \Upsilon \otimes_7 \mathcal{E}$$


---

**return**  $\langle \Upsilon, \mathcal{E} \rangle$

**Theorem 5.3.** Let  $\mathcal{G}$  be an instance of ST-REALIZABILITY.  $\mathcal{G}^* = \langle \Upsilon^*, \mathcal{E}^* \rangle$  can be computed using  $O(\log n)$  repeated applications of **SimpleSquare**( $\mathcal{G}$ ).

## 6 Parallel algorithms for SGSLogCFL

Let  $\mathcal{G} = \langle \Upsilon, \mathcal{E} \rangle$  be an instance of SYMMETRIC GAP UNDIRECTED ST-REALIZABILITY. Let the vertices of  $\mathcal{G}$  be  $V = \{1, 2, \dots, n\}$ .  $\mathcal{G}$  is represented by an  $n \times n$  standard matrix  $\mathcal{E}$  and an  $n^2 \times n^2$  gap matrix  $\Upsilon$ . In this section, we present parallel algorithms to compute  $\mathcal{G}$ 's transitive closure  $\mathcal{G}^* = \langle \Upsilon^*, \mathcal{E}^* \rangle$ . Let  $V^2 = V \times V$  be the set of pairs of vertices. In the rest of this paper the term ‘‘vertex’’ refers to elements from  $V$  as well as  $V^2$ . Let  $V^4 = V \times V \times V \times V$ .  $\mathcal{G}$  has two types of edges. The standard edges from  $V^2$  are present in  $\mathcal{E}$  and the gap edges from  $V^4$  are present in  $\Upsilon$ . In the rest of this paper the term ‘‘edge’’ refers to elements from  $V^2$  as well as  $V^4$ .

**Definition 6.1.** A subset of vertices  $S \subseteq V$  is a **standard component** ( $s$ -component) of  $\mathcal{G}$  iff for all  $u, v \in S$  it holds that  $(u \rightsquigarrow v)$  and  $(v \rightsquigarrow u)$ .

**Definition 6.2.** A subset  $S \subseteq V^2$  is a **gap component** ( $g$ -component) of  $\mathcal{G}$  iff for all  $(a, b), (c, d) \in S$  it holds that  $(a \rightsquigarrow (c, d) \rightsquigarrow b)$  and  $(c \rightsquigarrow (a, b) \rightsquigarrow d)$ .

In the rest of this paper the term ‘‘component’’ refers to both standard and gap components. If there is ambiguity we will explicitly say  $s$ -component or  $g$ -component.

A *pseudotree*  $P = (C, D)$  is a maximal connected directed graph with  $|C| = k$  vertices and  $|D| = k$  arcs for some  $k$ , for which each vertex has outdegree one. Note that every pseudotree has exactly one simple directed cycle (which may be a self-loop). The number of arcs in the cycle of a pseudotree  $P$  is its *circumference*. A *rooted tree* is a pseudotree whose cycle is a loop on some vertex  $r$  called the *root*. A *rooted star*  $R$  with root  $r$ , is a rooted tree whose arcs are of the form  $(x, r)$  with  $x \in R$ . A *pseudoforest* is a collection of pseudotrees.

**Symmetric Squaring :** We first present a simplified squaring algorithm when the input graph is an instance of SYMMETRIC GAP UNDIRECTED ST-REALIZABILITY. Note that the matrices  $\mathcal{E}$  and  $\Upsilon$  are symmetric i.e.,  $\mathcal{E}[a, b] = \mathcal{E}[b, a]$  and  $\Upsilon[(a, b), (c, d)] = \Upsilon[(c, d), (a, b)]$ . Moreover,  $\Upsilon[(a, b), (c, d)] = \Upsilon[(a, b), (d, c)] = \Upsilon[(b, a), (c, d)] = \Upsilon[(b, a), (d, c)]$ . Due to this symmetry, the products  $\otimes_3, \otimes_4, \otimes_6$  and  $\otimes_7$  are equivalent. Corollary 6.3 follows from Theorem 5.3.

---

**SymmetricSquare**( $\langle \Upsilon, \mathcal{E} \rangle$ )

$$\mathcal{E} = \mathcal{E} \otimes_1 \mathcal{E}$$

$$\mathcal{E} = \Upsilon \otimes_2 \mathcal{E}$$

$$\Upsilon = \Upsilon \otimes_3 \mathcal{E}$$

$$\Upsilon = \Upsilon \otimes_5 \Upsilon$$

**return**  $\langle \Upsilon, \mathcal{E} \rangle$

---

**Corollary 6.3.** Let  $\mathcal{G}$  be an instance of SYMMETRIC GAP UNDIRECTED ST-REALIZABILITY.  $\mathcal{G}^*$  can be computed using  $O(\log n)$  repeated applications of **SymmetricSquare**( $\mathcal{G}$ ).

## 6.1 An $O(\log^2 n)$ time parallel algorithm

---

**Connect**( $\mathcal{G} = \langle \Upsilon, \mathcal{E} \rangle$ )

```

1:  $\mathcal{E}^* \leftarrow \mathcal{E}$ 
2:  $\Upsilon^* \leftarrow \Upsilon$ 
3: for all  $i$  do  $X_{\mathcal{E}}(i) = i$ 
4: for all  $i$  do  $X_{\Upsilon}(i, j) = (i, j)$ 

5: for  $O(\log n)$  iterations do

6:   for all  $i$  do  $Temp_{\mathcal{E}}(i) \leftarrow StandardHook(i)$ 
7:   for all  $i$  do  $Temp_{\mathcal{E}}(i) \leftarrow \min_j \{Temp_{\mathcal{E}}(j) \mid X_{\mathcal{E}}(j) = i \text{ and } Temp_{\mathcal{E}}(j) \neq i\}$ 
8:   if none then  $Temp_{\mathcal{E}}(i) \leftarrow X_{\mathcal{E}}(i)$ 

9:   for all  $i$  do  $Temp_{\Upsilon}(i, j) \leftarrow GapHook(i, j)$ 
10:  for all  $i$  do  $Temp_{\Upsilon}(i, j) \leftarrow \min_{(k,l)} \{Temp_{\Upsilon}(k, l) \mid X_{\Upsilon}(k, l) = (i, j) \text{ and } Temp_{\Upsilon}(k, l) \neq (i, j)\}$ 
11:  if none then  $Temp_{\Upsilon}(i, j) \leftarrow X_{\Upsilon}(i, j)$ 

12:  for all  $i$  do  $X_{\mathcal{E}}(i) \leftarrow Temp_{\mathcal{E}}(i)$ 
13:  for all  $(i, j)$  do  $X_{\Upsilon}(i, j) \leftarrow Temp_{\Upsilon}(i, j)$ 

14:  for  $O(\log n)$  iterations do
15:    for all  $i$  do  $Temp_{\mathcal{E}}(i) \leftarrow Temp_{\mathcal{E}}(Temp_{\mathcal{E}}(i))$ 
16:    for all  $(i, j)$  do  $Temp_{\Upsilon}(i, j) \leftarrow Temp_{\Upsilon}(Temp_{\Upsilon}(i, j))$ 
17:  end for

18:  for all  $i$  do  $X_{\mathcal{E}}(i) \leftarrow \min\{Temp_{\mathcal{E}}(i), X_{\mathcal{E}}(Temp_{\mathcal{E}}(i))\}$ 
19:  for all  $(i, j)$  do  $X_{\Upsilon}(i, j) \leftarrow \min\{Temp_{\Upsilon}(i, j), X_{\Upsilon}(Temp_{\Upsilon}(i, j))\}$ 

20:  for all  $i, j$  do if  $X_{\mathcal{E}}(i) = X_{\mathcal{E}}(j)$  then  $\mathcal{E}^*[i, j] \leftarrow 1$ .
21:  for all  $i, j, k, l$  do if  $X_{\Upsilon}(i, j) = X_{\Upsilon}(k, l)$  then  $\Upsilon^*[i, (k, l), j] \leftarrow 1$ .

22: end for

23: return  $\mathcal{G}^* = \langle \Upsilon^*, \mathcal{E}^* \rangle$ 

```

---

We will assume that there is one processor  $P_i$  assigned to each vertex  $i \in V$ , one processor  $P_{ij}$  assigned to each edge  $(i, j) \in V^2$  and one processor  $P_{ijkl}$  assigned to each *gap edge*  $(i, j, k, l) \in V^4$ . We use a vector  $X_{\mathcal{E}}$  of length  $n$  to specify the  $s$ -components of  $\mathcal{G}$  as follows : if  $V_c \subseteq V$  is any  $s$ -component, then for all  $i \in V_c$ ,  $X_{\mathcal{E}}(i)$  equals the least element of  $V_c$ . We use an  $n \times n$  matrix  $X_{\Upsilon}$  to specify the  $g$ -components of  $\mathcal{G}$  as follows : if  $W_c \subseteq V^2$  is any  $g$ -component, then for all  $(i, j) \in W_c$ ,  $X_{\Upsilon}(i, j)$  equals the lexicographically least element of  $W_c$ .

The algorithm **Connect** iteratively computes the vectors  $X_{\mathcal{E}}$  and  $X_{\Upsilon}$  from the input  $\mathcal{G} = \langle \Upsilon, \mathcal{E} \rangle$  and updates  $\Upsilon^*$  and  $\mathcal{E}^*$ . It is based on a hook and contract algorithm [HCS79] that works as follows. The algorithm deals with “components”, which are sets of “vertices” *found* to belong to the same (standard or gap) component of  $\mathcal{G}$ . Each component is equipped with an edge-list, a linked list of edges that connect

---

**StandardHook( $i$ )**

```
1:  $S_1 \leftarrow \{X_{\mathcal{E}}(j) \mid \mathcal{E}^*[i, j] = 1 \text{ and } X_{\mathcal{E}}(j) \neq X_{\mathcal{E}}(i)\}$ 
2:  $S_2 \leftarrow \{X_{\mathcal{E}}(j) \mid \Upsilon^*[i, (k, k), j] = 1 \text{ and } X_{\mathcal{E}}(j) \neq X_{\mathcal{E}}(i)\}$ 
3:  $S = S_1 \cup S_2$ 
4: if  $S = \emptyset$  then
5:   return  $X_{\mathcal{E}}(i)$ 
6: else
7:   return  $\min(S)$ 
8: end if
```

---

---

**GapHook( $i, j$ )**

```
1:  $S_1 \leftarrow \{X_{\Upsilon}(k, l) \mid \Upsilon^*[i, (k, l), j] = 1 \text{ and } X_{\Upsilon}(k, l) \neq X_{\mathcal{E}}(i, j)\}$ 
2:  $S_2 \leftarrow \{X_{\Upsilon}(k, j) \mid \mathcal{E}^*[i, k] = 1 \text{ and } X_{\Upsilon}(k, j) \neq X_{\Upsilon}(i, j)\}$ 
3:  $S = S_1 \cup S_2$ 
4: if  $S = \emptyset$  then
5:   return  $X_{\Upsilon}(i, j)$ 
6: else
7:   return  $\min(S)$ 
8: end if
```

---

it to other components. Initially each element from  $V$  is an  $s$ -component by itself. Their edge-lists correspond to the undirected edges of  $\mathcal{E}$ . These components will eventually grow and become the corresponding  $s$ -components. Initially each element from  $V^2$  is a  $g$ -component by itself. Their edge-lists correspond to the undirected edges of  $\Upsilon$ . These components will eventually grow and become the corresponding  $g$ -components. The algorithm proceeds as follows :

**repeat** until there are no edges left :

1. Each component picks an edge pointing to a lexicographically minimum vertex from its edge-list leading to a neighboring component and hooks by pointing to it. If a component has an empty edge-list, it hooks to itself. The details of hooking are presented in **StandardHook** and **GapHook**. Note that both these hooking steps use the previously computed connectivity information from *both*  $\Upsilon^*$  and  $\mathcal{E}^*$ . These hooking processes create clusters of components called pseudotrees. The  $s$ -components form pseudotrees on the vertex set  $V$  and  $g$ -components form pseudotrees on the vertex set  $V^2$ .
2. Each pseudotree is identified as a new component with one of its vertices as its representative. Each representative receives into its edge-list all the edges contained in the edge-lists of its pseudotree. At this stage the matrices  $\mathcal{E}^*$  and  $\Upsilon^*$  are updated with “new” edges.
3. Edges internal to components are removed implicitly.

During the first iteration the edges connecting each vertex to neighboring vertices are examined (steps 6-11), and sets of vertices which are known to be connected are identified (steps 14-17). In effect, each such set of vertices is merged into a “supervertex” which are specified by the vectors  $X_{\mathcal{E}}(i)$  and  $X_{\Upsilon}(i, j)$ . For each  $i$  in a supervertex,  $X_{\mathcal{E}}(i)$  equals the smallest-numbered vertex in the supervertex. For each  $(i, j)$  in a supervertex,  $X_{\Upsilon}(i, j)$  equals the lexicographically first vertex in the supervertex. In succeeding iterations,



the edges connecting each supervertex to neighboring supervertices are examined in steps 6-11, and sets of supervertices are merged in steps 14-17. The process continues until all the vertices in a (standard and gap) component have been merged into one gigantic supervertex.

**Theorem 6.4.** The algorithm **Connect** finds  $\mathcal{G}^* = \langle \Upsilon^*, \mathcal{E}^* \rangle$  in parallel time  $O(\log^2 n)$  using  $n^4$  processors in the CREW PRAM model.

**Connect** algorithm is a generalization of the parallel algorithm presented in [HCS79]. We added two hooking procedures (one for growing  $s$ -components and one for growing  $g$ -components). Unlike [HCS79] the new edges found after contraction step are added in the matrices  $\Upsilon^*$  and  $\mathcal{E}^*$  *before* starting the next hooking step.

The algorithms of [JM97] and [CL95] can similarly be generalized to compute  $\mathcal{G}^* = \langle \Upsilon^*, \mathcal{E}^* \rangle$  in parallel time  $O(\log^{3/2} n)$  and  $O(\log n \log \log n)$  respectively. The processor bounds in all these algorithms is polynomial in  $n$ , the number of vertices of  $\mathcal{G}$ . We now present an outline of the parallel algorithms of [JM97] and [CL95] and the necessary modifications to apply them to **SGSLogCFL**.

## 6.2 An $O(\log^{3/2} n)$ time parallel algorithm

In the algorithm presented in the previous section the size of the components formed after hooking phase may vary a lot. A slow growing component may consist of as few as two vertices, whereas a fast-growing component may have as many as  $n$  vertices for an  $s$ -component and  $n^2$  vertices for a  $g$ -component. As a result the contraction (steps 14-17) requires  $\Theta(\log n)$  time in order to allow the biggest component to contract to a single vertex. The algorithm must iterate  $\log n$  times so that a slow-growing component, which may only double its size in each iteration, can eventually grow to its full size. An crucial observation of [JM97] is that slow-growing components need little time to contract and fast-growing components require fewer iterations to grow to their full size.

Johnson and Metaxas [JM97] presented an algorithm in which components are scheduled to hook and contract according to their growth rate. Their algorithm schedules every component to grow by a factor of at least  $2^{\sqrt{\log n}}$  in a phase of  $O(\log n)$  time. Hence,  $\sqrt{\log n}$  phases suffice to find all connected components in the graph, for a total of  $O(\log^{3/2} n)$  time. Within a phase slow-growing components are scheduled to hook and contract in  $o(\log n)$  time repeatedly until they catch up with fast-growing components. Fast-growing components are left idle once they have achieved the intended size.

- In the algorithm of [HCS79] the vertices hook to a lexicographically minimum vertex. In Johnson-Metaxas algorithm vertices hook to the *first edge* in their edge-list. This creates pseudotrees of arbitrary circumference i.e., pseudotrees can have large cycles which are to be contracted properly in the contraction phase. Since exclusive writing is required, the usual pointer doubling technique will not terminate when applied to a cycle. Johnson and Metaxas [JM97] introduced *cycle-reducing shortcutting* technique to solve this problem. This technique (i) contracts a pseudotree into a rooted tree in time logarithmic in its circumference, (ii) contracts a rooted tree into a rooted star in time logarithmic in the length of its longest path.
- It is expensive to compute the set of edges of all the components in a pseudotree without concurrent writing. Potentially there are a large number of components that hook together in the first step and therefore a large number of components that are ready to give their edge-lists simultaneously to the new super-component's edge-list. Johnson and Metaxas [JM97] introduced *edge-plugging scheme* which achieves the objective in constant time, irrespective of whether the component is yet contracted to a rooted star.

- It is also expensive to have a component pick a mate. There may be a large number of edges internal to the component. The number of such edges grows every time components hook. These internal edges cannot be used to find a mate. Hence, a component may attempt to find a mate several times and unsuccessful if it picks an internal edge. Removing all the internal edges before picking an edge may also take a lot of time. Johnson and Metaxas [JM97] introduced a *growth-control schedule*. Components grow in size in a uniform way that controls their minimum sizes as long as continued growth is possible. The internal edges are identified and removed periodically to make hooking more efficient. The algorithm recognizes whether a component is growing too fast and therefore can be ignored.

For implementation details of the above algorithm see [JM97]. As mentioned earlier, to get the corresponding parallel algorithm for **SGSLogCFL** we add two hooking procedures (one for growing  $s$ -components and one for growing  $g$ -components). After each contraction step the newly found edges are added in the matrices  $\Upsilon^*$  and  $\mathcal{E}^*$ .

### 6.3 An $O(\log n \log \log n)$ time parallel algorithm

The Chong Lam algorithm [CL95] is also based on a hook and contract approach. The hooking process uses an ordering  $<_d$  of the vertices such that  $u <_d v$  iff the degree of  $u$  is less than the degree of  $v$  (or the degrees are the same, but  $u$  is less than  $v$  in their lexicographic ordering). Before every phase, every vertex of the current *supergraph* is either active, inactive or done. All active and inactive vertices have nonzero degree, the done vertices have zero degree, and there are no multiedges between active vertices; the inactive vertices are organized in a set of hooking trees. Initially all vertices with nonzero degree are active, and the rest are done.

To choose their hooking edges, the active vertices of the graph perform the following steps in parallel.  
 (i) if a vertex  $v$  has a neighbor larger according to  $<_d$  than itself, then  $v$  hooks to the *largest* such neighbor.  
 (ii) if after the first step all neighbors of  $v$  are hooked to it, then  $v$  hooks to itself. Otherwise, if after the first step a neighbor  $u$  of  $v$  is hooked to a vertex different from  $v$ , then  $v$  hooks to  $u$ .

Some of the current hooking trees are contracted to a representative vertex in a contraction phase. The representative vertex is the only vertex in the tree which is hooked to itself. Whether a tree is contracted is determined by a parameter. This parameter depends on the phase and sets an upper bound on the sum of the degrees of the vertices of the trees which are contracted. For every contracted tree, its representative becomes a new active vertex and the rest of its vertices become done. All multiedges between new active vertices are removed. The vertices of every uncontracted tree become inactive.

The processing required by a hooking phase is performed in parallel time  $O(\log d)$ , where  $d$  is the degree of the active vertex, using pointer jumping. Checking the degree of a hooking tree during the contraction phase is done in parallel time  $O(\log c)$ , where  $c$  is the contraction parameter, by using pointer jumping and a constant time edge-list plugging technique.

---

**Connect( $k$ )**

**MaxHook;**

**if**  $k > 0$  **then**

**Connect**( $2^{2^k}$ )

**Connect**( $k - 1$ )

**Connect**( $k - 1$ )

## Contract( $2^{2^{k+1}}$ )

---

A call to **Connect**( $\lceil \log \log n \rceil$ ) contracts every connected component of the graph to a single vertex and all the other vertices are organized in a set of rooted parent trees such that the root of the tree of a vertex  $u$  is the vertex to which the connected component of  $u$  is contracted.

To generalize this algorithm to **SGSLogCFL**, we make the following modifications : (i) add two hooking procedures (one for growing  $s$ -components and one for growing  $g$ -components) (ii) the new edges found after every call *Contract* are added in the matrices  $\Upsilon^*$  and  $\mathcal{E}^*$  and the new degrees of the vertices are recomputed. The correctness of the algorithm follows by using Corollary 6.3 in the correctness argument of [CL95], implying an  $O(\log n \log \log n)$  time EREW parallel algorithm computing  $\mathcal{G}^* = \langle \Upsilon^*, \mathcal{E}^* \rangle$ .

## 7 **SGSLogCFL** $\subseteq$ *DSPACE*( $\log n \log \log n$ )

In this section we present an outline of Trifonov's  $O(\log n \log \log n)$ -space algorithm for **UNDIRECTED ST-CONNECTIVITY** [Tri08]. Trifonov's algorithm is based on the  $O(\log n \log \log n)$  time deterministic EREW PRAM algorithm with  $O(m + n)$  processors of Chong and Lam [CL95] outlined in the previous section. This parallel algorithm is first simulated sequentially in linear space. Using this sequential algorithm a mathematical structure called *configuration* is defined. This configuration corresponds to the state of the sequential algorithm at a certain point of its execution. An ordering on the edges incident to a vertex is fixed, and the hooking is done sequentially for all active vertices. Using the sequence of configurations an  $O(\log^2 n)$  space algorithm, which instead of storing all of its current state recomputes parts of it when it needs them. This algorithm works pretty much like Savitch's algorithm [Sav70].

The max-degree hooking scheme of [CL95] ensures that small trees have small neighborhoods. Using the exploration walks on trees defined by Koucky [Kou02], the levels of recursion of [CL95] are implemented so that they process small trees in  $o(\log n)$  space. These walks essentially play the role of the edge-list plugging technique and pointer jumping techniques employed by the Chong-Lam algorithm. They allow us to traverse trees space-efficiently.

The  $O(\log n)$  space per level is mainly due to storing vertices in the local variables of the functions, since each vertex takes  $\Theta(\log n)$  space. To overcome this bottleneck the functions are redefined so that they never keep a vertex in their local variables. The vertex  $v$  is removed from the argument list of the functions. Instead of this argument, one current vertex is maintained in a global variable. All functions are programmed to return some "information" about this vertex. A function which otherwise must return a vertex is defined so that after its execution the current vertex is its result. If needed the calling function keeps enough information locally to restore the original current vertex. The crucial part of the optimization is to avoid storing vertices locally and be able to move the current vertex temporarily, perform something at the new current vertex, and then return to the original current vertex. Instead of this going back and forth between the two vertices, using the reversibility of the moves along the edges and the exploration walks on the trees, the comparison is performed bit by bit. Aside from the information stored for the ways back, this takes only the  $\Theta(\log n \log \log n)$  space necessary to store the index of a bit. In this way the bottleneck of  $\Omega(\log n)$  space is reduced to  $\Omega(\log n \log \log n)$ . The introduction of one global current vertex and always returning information about this vertex, mimics the implementation and correctness of Chong-Lam algorithm with minor modifications to the hooking scheme. The current vertex is an implicit argument to all functions describing a *configuration*.

To generalize this algorithm to **SGSLogCFL**, we make the following modifications : (i) add two

hooking procedures (one for growing  $s$ -components and one for growing  $g$ -components) (ii) the new edges found after every call *Contract* are added in the matrices  $\Upsilon^*$  and  $\mathcal{E}^*$  and the new degrees of the vertices are recomputed and (iii) the exploration walks and the bit by bit comparison are done on the hooking trees generated by  $s$ -components and  $g$ -components.

**Theorem 7.1.** Let  $\mathcal{G} = \langle \Upsilon, \mathcal{E} \rangle$  be an instance of SYMMETRIC GAP UNDIRECTED ST-REALIZABILITY.  $\mathcal{G}^* = \langle \Upsilon^*, \mathcal{E}^* \rangle$  can be computed in  $O(\log n \log \log n)$  space i.e.,  $\text{SGSLogCFL} \in \text{DSPACE}(\log n \log \log n)$ .

**Corollary 7.2.**  $\text{BALANCED ST-CONNECTIVITY} \in \text{DSPACE}(\log n \log \log n)$ .

## 8 Open Problems

In a recent work [Kin10], we proved that  $\text{BALANCED ST-CONNECTIVITY}$ ,  $\text{SGSLogCFL}$  and  $\text{POSITIVE BALANCED ST-CONNECTIVITY}$  are all closed under complementation. Several interesting research directions arise from our work :

- **Balanced Connectivity:**  $\text{BALANCED ST-CONNECTIVITY}$  and  $\text{POSITIVE BALANCED ST-CONNECTIVITY}$  are natural graph connectivity problems that lie between  $\mathbf{L}$  and  $\mathbf{NL}$ . Studying their space complexity is an interesting research direction towards improving the space complexity of  $\text{ST-CONNECTIVITY}$ . In particular, it would be interesting to improve Theorem 7.1. Is  $\text{BALANCED ST-CONNECTIVITY} \in \mathbf{L}$  ? Less ambitiously, is  $\text{BALANCED ST-CONNECTIVITY} \in \mathbf{SC}^2$  ?
- An alternate proof of Theorem 7.1 using the techniques of [RVW02, Rei08] or [RV05] seems to be a challenging task.
- **SLogCFL vs LogDCFL:** In the logspace setting we have  $\mathbf{L} = \mathbf{SL} \subseteq \mathbf{NL}$ . In the  $\text{LogCFL}$  setting, we have  $\text{LogDCFL} \subseteq \text{SLogCFL} = \text{LogCFL}$  (see Theorem 4.2). By definition, we have  $\mathbf{NL} \subseteq \text{LogCFL}$ . It is known that  $\text{LogDCFL} \subseteq \mathbf{SC}^2$  [Coo79]. This motivates the study of the relationship between  $\text{LogDCFL}$  and  $\text{SLogCFL}$ . It would be interesting to generalize the techniques of [RVW02, Rei08] to prove  $\text{LogDCFL} = \text{SLogCFL}$ . This would imply  $\mathbf{NL} \subseteq \mathbf{SC}^2$ , i.e.,  $\text{ST-CONNECTIVITY}$  can be solved by a deterministic algorithm in polynomial time and  $O(\log^2 n)$  space.
- **SLogCFL vs RLogCFL:** We have  $\text{LogDCFL} \subseteq \text{SLogCFL} = \text{LogCFL}$  and  $\text{LogDCFL} \subseteq \text{RLogCFL} \subseteq \text{LogCFL}$  implying  $\text{RLogCFL} \subseteq \text{SLogCFL}$ . In the logspace setting, prior to Reingold's work, Aleliunas et. al. [AKL<sup>+</sup>79] proved that  $\mathbf{SL} \subseteq \mathbf{RL}$ , using random walks. It would be interesting to generalize their techniques to prove  $\text{SLogCFL} \subseteq \text{RLogCFL}$ . Since  $\text{BPLogCFL} \subseteq \mathbf{SC}^2$  [Ven06], a proof of  $\text{SLogCFL} \subseteq \text{RLogCFL}$  would imply  $\mathbf{NL} \subseteq \mathbf{SC}^2$ .
- Is there a circuit characterization of  $\text{SGSLogCFL}$  ? What is the relationship between (i)  $\text{SGSLogCFL}$  and  $\mathbf{NL}$  ? (ii)  $\text{SGSLogCFL}$  and  $\text{LogDCFL}$  ? (iii)  $\text{SGSLogCFL}$  and  $\text{DET}^1$  ?
- Allender and Lange [AL10] proved that  $\text{SLogCFL} = \text{LogCFL}$ . Is  $1\text{SLogCFL} = 1\text{LogCFL}$  ? i.e., is  $\text{POSITIVE BALANCED ST-CONNECTIVITY}$   $\mathbf{NL}$ -complete ?

---

<sup>1</sup> $\text{DET}$  is the class of problems  $\text{NC}^1$  Turing reducible to the determinant [Coo85].

**Acknowledgements** : This project is partially funded by the NSF grant CCF-0902717. I gratefully acknowledge helpful discussions with Eric Allender, Klaus-Jörn Lange, Nutan Limaye, Richard J. Lipton, H. Venkateswaran and Dieter van Melkebeek.

## References

- [AKL<sup>+</sup>79] Romas Aleliunas, Richard M. Karp, Richard J. Lipton, László Lovász, and Charles Rackoff. Random walks, universal traversal sequences, and the complexity of maze problems. *FOCS*, pages 218–223, 1979.
- [AL10] Eric Allender and Klaus-Jörn Lange. Symmetry coincides with nondeterminism for time-bounded auxiliary pushdown automata. *To appear in 25th Computational Complexity Conference*, 2010.
- [All] Eric Allender. Personal communication.
- [All07] Eric Allender. Reachability problems: An update. In *CiE*, pages 25–27, 2007.
- [AP87] Foto N. Afrati and Christos H. Papadimitriou. The parallel complexity of simple chain queries. In *PODS*, pages 210–213, 1987.
- [ATWZ00] Roy Armoni, Amnon Ta-Shma, Avi Wigderson, and Shiyu Zhou. An  $(\log^{4/3})$  space algorithm for  $(s, t)$  connectivity in undirected graphs. *J. ACM*, 47(2):294–311, 2000.
- [BCD<sup>+</sup>89] Allan Borodin, Stephen A. Cook, Patrick W. Dymond, Walter L. Ruzzo, and Martin Tompa. Two applications of inductive counting for complementation problems. *SIAM J. Comput.*, 18(3):559–578, 1989.
- [Bor77] Allan Borodin. On relating time and space to size and depth. *SIAM J. Comput.*, 6(4):733–744, 1977.
- [CHL99] Ka Wong Chong, Yijie Han, and Tak Wah Lam. On the parallel time complexity of undirected connectivity and minimum spanning trees. In *SODA*, pages 225–234, 1999.
- [CL95] Ka Wong Chong and Tak Wah Lam. Finding connected components in  $o(\log n \log \log n)$  time on the erew pram. *J. Algorithms*, 18(3):378–402, 1995.
- [Coo71] Stephen A. Cook. Characterizations of pushdown machines in terms of time-bounded computers. *J. ACM*, 18(1):4–18, 1971.
- [Coo79] Stephen A. Cook. Deterministic CFL’s are accepted simultaneously in polynomial time and log squared space. In *STOC*, pages 338–345, 1979.
- [Coo85] Stephen A. Cook. A taxonomy of problems with fast parallel algorithms. *Information and Control*, 64(1-3):2–21, 1985.
- [GHR95] Raymond Greenlaw, H. James Hoover, and Walter L. Ruzzo. Limits to parallel computation: P-completeness theory. *Oxford University Press*, 1995.
- [Gre73] Sheila A. Greibach. The hardest context-free language. *SIAM J. Comput.*, 2(4):304–310, 1973.

- [GW96] Anna Gál and Avi Wigderson. Boolean complexity classes vs. their arithmetic analogs. *Random Struct. Algorithms*, 9(1-2):99–111, 1996.
- [Har78] Michael A. Harrison. Introduction to formal languages theory. *Addison-Wesley series in computer science*, 1978.
- [HCS79] Daniel S. Hirschberg, Ashok K. Chandra, and Dilip V. Sarwate. Computing connected components on parallel computers. *Commun. ACM*, 22(8):461–464, 1979.
- [Hen77] F. Hennie. Introduction to computability. *Addison-Wesley, Reading, MA*, 1977.
- [Imm88] Neil Immerman. Nondeterministic space is closed under complementation. *SIAM J. Comput.*, 17:935–938, 1988.
- [JM95] Donald B. Johnson and Panagiotis Takis Metaxas. A parallel algorithm for computing minimum spanning trees. *J. Algorithms*, 19(3):383–401, 1995.
- [JM97] Donald B. Johnson and Panagiotis Takis Metaxas. Connected components in  $o(\log^{3/2}n)$  parallel time for the crew pram. *J. Comput. Syst. Sci.*, 54(2):227–242, 1997.
- [Kin10] Shiva Kintali. Realizable Paths and the Closure Under Complementation. *Under Preparation*, 2010.
- [Kou02] Michal Koucký. Universal traversal sequences with backtracking. *J. Comput. Syst. Sci.*, 65(4):717–726, 2002.
- [KW93] Mauricio Karchmer and Avi Wigderson. On span programs. In *Structure in Complexity Theory Conference*, pages 102–111, 1993.
- [Lim05] Nutan Limaye. Parallel complexity classes centered around LogCFL. *M.Sc. thesis, Anna University*, 2005.
- [LP82] Harry R. Lewis and Christos H. Papadimitriou. Symmetric space-bounded computation. *Theoretical Computer Science*, 19:161–187, 1982.
- [MR00] David Melski and Thomas W. Reps. Interconvertibility of a class of set constraints and context-free-language reachability. *Theoretical Computer Science*, 248(1-2):29–98, 2000.
- [MRV99] Pierre McKenzie, Klaus Reinhardt, and V. Vinay. Circuits and context-free languages. In *COCOON*, pages 194–203, 1999.
- [Nis94] Noam Nisan.  $RL \subseteq SC$ . *Computational Complexity*, 4:1–11, 1994.
- [NR95] Rolf Niedermeier and Peter Rossmanith. Unambiguous auxiliary pushdown automata and semi-unbounded fan-in circuits. *Information and Computation*, 118(2):227–245, 1995.
- [NSW92] Noam Nisan, Endre Szemerédi, and Avi Wigderson. Undirected connectivity in  $O(\log^{3/2}n)$  space. In *FOCS*, pages 24–29, 1992.
- [NT95] Noam Nisan and Amnon Ta-Shma. Symmetric logspace is closed under complement. *STOC*, pages 140–146, 1995.

- [Rei08] Omer Reingold. Undirected connectivity in log-space. *J. ACM*, 55(4), 2008.
- [Rep96] Thomas W. Reps. On the sequential nature of interprocedural program-analysis problems. *Acta Inf.*, 33(8):739–757, 1996.
- [Ruz80] Walter L. Ruzzo. Tree-size bounded alternation. *J. Comput. Syst. Sci.*, 21(2):218–235, 1980.
- [RV05] Eyal Rozenman and Salil P. Vadhan. Derandomized squaring of graphs. In *APPROX-RANDOM*, pages 436–447, 2005.
- [RVW00] Omer Reingold, Salil P. Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. In *FOCS*, pages 3–13, 2000.
- [RVW02] Omer Reingold, Salil P. Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. *Annals of Mathematics*, 155(1), 2002.
- [Sav70] Walter J. Savitch. Relationships between nondeterministic and deterministic tape complexities. *J. Comput. Syst. Sci.*, 4(2):177–192, 1970.
- [SBV10] Derrick Stolee, Chris Bourke, and N. V. Vinodchandran. A log-space algorithm for reachability in planar acyclic digraphs with few sources. In *IEEE Conference on Computational Complexity*, pages 131–138, 2010.
- [Sud78] Ivan Hal Sudborough. On the tape complexity of deterministic context-free languages. *J. ACM*, 25(3):405–414, 1978.
- [SV10] Derrick Stolee and N. V. Vinodchandran. Space-efficient algorithms for reachability in surface-embedded graphs. *Electronic Colloquium on Computational Complexity*, TR10-154, 2010.
- [SZ99] Michael E. Saks and Shiyu Zhou.  $BP_HSPACE(S) \subseteq DSPACE(S^{3/2})$ . *Journal of Computer and System Sciences*, 58(2):376–403, 1999.
- [Sze87] Róbert Szelepcsényi. The method of forcing for nondeterministic automata. *Bulletin of EATCS*, 33:96–100, 1987.
- [Tri08] Vladimir Trifonov. An  $O(\log n \log \log n)$  space algorithm for undirected st-connectivity. *SIAM J. Comput.*, 38(2):449–483, 2008.
- [UG86] Jeffrey D. Ullman and Allen Van Gelder. Parallel complexity of logical query programs. In *FOCS*, pages 438–454, 1986.
- [Ven91] H. Venkateswaran. Properties that characterize LogCFL. *J. Comput. Syst. Sci.*, 43(2):380–404, 1991.
- [Ven06] H. Venkateswaran. Derandomization of probabilistic auxiliary pushdown automata classes. *IEEE Conference on Computational Complexity*, pages 355–370, 2006.
- [Wig94] Avi Wigderson.  $NL/poly \subseteq \oplus L/poly$  (preliminary version). In *Structure in Complexity Theory Conference*, pages 59–62, 1994.

# Appendix

## A Symmetric AuxPDAs

An **auxiliary pushdown automaton** (AuxPDA) is a multi-tape Turing machine with a two-way read-only input tape, a pushdown tape, and one or more work tapes. The pushdown alphabet has a distinguished symbol (say  $\$$ ) which is initially *pushed* on the pushdown tape. The machine is designed so that the pushdown head never shifts left of  $\$$  or changes  $\$$ . Further, the pushdown head can never shift left when scanning any tape symbol unless it first erases (i.e., pops) that symbol, and it can never shift right from a square unless it first prints (i.e., pushes) a nonblank symbol on that square. Space on an AuxPDA is the space used on the work tapes without counting the space on the pushdown tape. Formally, an AuxPDA is an 8-tuple  $M = (Q, \Sigma, \Sigma_0, \Sigma_\alpha, l, \Delta, s, F)$ , where  $Q$  is a finite set of states,  $\Sigma$  is a finite tape alphabet,  $\Sigma_0 \subseteq \Sigma$  is the input alphabet,  $\Sigma_\alpha \subseteq \Sigma$  is the pushdown alphabet,  $l$  is the number of tapes,  $s \in Q$  is the initial state,  $F \subseteq Q$  is the set of final states and  $\Delta$  is a finite set of transitions.

We first define the *transition* of an AuxPDA that enable the AuxPDA to “peek” one square right or left on the input and work tapes and one square below the top symbol of the pushdown tape while changing its configuration. A transition is of the form  $(p, \mathcal{S}, t_1, \dots, t_l, q)$ , where  $p$  and  $q$  are states,  $\mathcal{S}$  is a stack triple,  $l$  is the number of tapes, and  $t_1, \dots, t_l$  are tape triples. A stack triple is either of the form (i)  $(\alpha_a \alpha_b, P, \alpha_c \alpha_d)$ , where  $\alpha_a, \alpha_b, \alpha_c, \alpha_d \in \Sigma_\alpha$  and  $P$  is  $+1$  or  $-1$ ; or is of the form (ii)  $(\alpha_a, 0, \alpha_b)$ , where  $\alpha_a, \alpha_b \in \Sigma_\alpha$ . A tape triple is either of the form (i)  $(ab, D, cd)$ , where  $a, b, c, d \in \Sigma$  and  $D$  is  $+1$  or  $-1$ ; or is of the form (ii)  $(a, 0, b)$ , where  $a, b \in \Sigma$ .

A transition of the form  $(p, \mathcal{S}, t_1, \dots, t_l, q)$  signifies that  $M$  moves from state  $p$  to state  $q$  according to the stack and tape triples. The tape triple  $t_i = (ab, +1, cd)$  signifies that when  $M$  is scanning symbol  $a$  on tape  $t_i$ , and with the square just to the right of the scanned square containing symbol  $b$ ,  $M$  may rewrite these two squares to contain symbols  $c$  and  $d$ , respectively, move its tape head one square to the right. Similarly, a transition  $(ab, -1, cd)$  signifies a potential left movement of the tape head, except that now the scanned symbol must be  $b$  and the one to its left  $a$  and these are rewritten as  $d$  and  $c$ , respectively. The tape triple  $(a, 0, b)$  signifies that  $M$  replaces the symbol  $a$  with  $b$  without moving its head position. The stack triple is defined analogously with  $P = +1$  (resp.  $P = -1$ ) corresponding to a push (resp. pop) operation on the pushdown tape.

The **surface configuration** (introduced by Cook [Coo71]) of an AuxPDA on an input  $w$  consists of the state, contents and head positions of the work tapes, the head position of the input tape and the topmost symbol of the stack. Note that for a space  $S(n)$ -bounded AuxPDA, its surface configurations take only  $O(S(n))$  space. In the rest of this section, we will refer to surface configurations as configurations. Let  $\mathcal{C}(M)$  denote the set of all configurations of  $M$ . For an input  $w$ , and  $C_1, C_2 \in \mathcal{C}(M)$  we write  $C_1 \vdash_M C_2$  to denote that  $C_1$  “yields”  $C_2$ . A *computation* by  $M$  is a sequence  $C_0 \vdash_M C_1 \vdash_M \dots \vdash_M C_n$ , where  $n \geq 0$  and  $C_0, \dots, C_n \in \mathcal{C}(M)$ . The reflexive, transitive closure of  $\vdash_M$  is denoted by  $\vdash_M^*$  and the transitive closure is denoted by  $\vdash_M^+$ . An AuxPDA  $M$  is nondeterministic (resp. deterministic) if  $\vdash_M$  is multi-valued (resp. single-valued).

Since the tape triples and stack triples of  $M$  enable it to peek into only a *constant* number of symbols,  $M$  can be simulated by a standard AuxPDA extending the notion of *big-headed* Turing machines [Hen77]. The “peeking” version of  $M$  enables us to define symmetric computation. Each transition  $\delta = (p, \mathcal{S}, t_1, \dots, t_l, q)$  has an inverse  $\delta^{-1} = (q, \mathcal{S}^{-1}, t_1^{-1}, \dots, t_l^{-1}, p)$  where if  $\mathcal{S} = (\alpha, P, \beta)$  then  $\mathcal{S}^{-1} = (\beta, -P, \alpha)$  and for  $i = 1, \dots, k$  if  $t_i = (a, D, b)$  then  $t_i^{-1} = (b, -D, a)$ .

The inverse of an AuxPDA  $M = (Q, \Sigma, \Sigma_0, \Sigma_\alpha, l, \Delta, s, F)$  is  $M^{-1} = (Q, \Sigma, \Sigma_0, \Sigma_\alpha, l, \Delta^{-1}, s, F)$ ,



where  $\Delta^{-1} = \{\delta^{-1} : \delta \in \Delta\}$ . An AuxPDA is *symmetric* if it is its own inverse i.e., if  $\delta^{-1} \in \Delta$  whenever  $\delta \in \Delta$ . The symmetric closure of an AuxPDA  $M = (Q, \Sigma, \Sigma_0, \Sigma_\alpha, l, \Delta, s, F)$  is  $\overline{M} = (Q, \Sigma, \Sigma_0, \Sigma_\alpha, l, \Delta \cup \Delta^{-1}, s, F)$ . Note that the symmetric closure of an AuxPDA is symmetric and a symmetric AuxPDA is its own symmetric closure. We now define the complexity class **SLogCFL**.

**SLogCFL** is the class of languages accepted by log space bounded and polynomial time bounded symmetric AuxPDA.

Let  $\#$  be a new special symbol in the tape alphabet that does not belong to input alphabet. For an AuxPDA  $M$ ,  $M^\#$  is its normal form such that (1)  $M$  and  $M^\#$  accept the same language in the same space bound, and have the same number of tapes; (2)  $M^\#$  has no transitions into its initial state or out of any final state; (3) for any configurations  $C_1, C_2 \in \mathcal{C}(M^\#)$  if  $C_1 \vdash_{M^\#} C_2$  then  $|C_1| \leq |C_2|$ , where  $|C|$  represents the space of  $C$ .  $M^\#$  is constructed from  $M$  by adding a new initial state and transitions from it to the old initial state; eliminating any transitions out of final states; and introducing a new *pseudoblank* symbol which  $M^\#$  writes instead of writing (or rewriting) a blank on a worktape, and which  $M^\#$  treats as indistinguishable from a blank when seen on a worktape.  $\overline{M^\#}$  is the symmetric closure of  $M^\#$ .

The following lemma is proved by Lewis and Papadimitriou [LP82] in the context of symmetric Turing machines. By our definition of symmetric AuxPDA's, its proof follows by treating the ‘‘configurations’’ of a symmetric Turing machine as the ‘‘surface configurations’’ of a symmetric AuxPDA and augmenting the transitions with stack triples. We skip its proof since it is essentially the proof of [LP82].

**Lemma A.1.** Let  $M = (Q, \Sigma, \Sigma_0, \Sigma_\alpha, l, \Delta, s, F)$  be any AuxPDA, and let  $\mathcal{A} \subseteq \mathcal{C}(M)$ . Suppose that

- (a) for any  $A_1, A_2 \in \mathcal{A}$ , if  $A_1 \vdash_M^{+\mathcal{A}} A_2$  then  $A_2 \vdash_M^{+\mathcal{A}} A_1$
- (b) for any  $A \in \mathcal{A} \cup \mathcal{I}(M)$ , and  $B \notin \mathcal{A}$  and any  $C_1, C_2, C_3$ , if  $A \vdash_M^{*\mathcal{A}} C_1 \dashv_M^{*\mathcal{A}} C_2 \dashv_M B \vdash_M C_3$ , then  $C_2 = C_3$
- (c) for any  $A_1 \in \mathcal{A} \cup \mathcal{I}(M)$ , any  $A_2 \in \mathcal{A}$ , and any  $B$ , if  $A_1 \vdash_M^{*\mathcal{A}} B \dashv_M^{*\mathcal{A}} A_2$  then  $A_1 = A_2$ .

Then  $\overline{M^\#}$  accepts the same language as  $M$  in the same space as  $M$ .

**Theorem 3.1**  $\mathbf{LogDCFL} \subseteq \mathbf{SLogCFL} \subseteq \mathbf{LogCFL}$ .

*Proof.* Let  $M$  be a deterministic logspace bounded AuxPDA accepting a language  $L \in \mathbf{LogDCFL}$ . Then  $M$  satisfies the hypothesis of Lemma A.1, with  $\mathcal{A} = \emptyset$ .  $\overline{M}$  satisfies the hypothesis (a) and (c) trivially. Since  $M$  is deterministic it satisfies the hypothesis (b). Hence  $\overline{M^\#}$  accepts  $L$ . Hence,  $\mathbf{LogDCFL} \subseteq \mathbf{SLogCFL}$ . The second inclusion is trivial, since nondeterminism is more general than symmetry.  $\square$

Now that we have the definition and properties of **SLogCFL**, the proofs of the following theorem and its corollary are similar to those of Theorem 2.2 and Corollary 2.3. It is routine to check that the AuxPDA thus constructed, satisfies the properties of Lemma A.1.

**Theorem 3.2** **UNDIRECTED ST-REALIZABILITY** is **SLogCFL**-complete.

**Corollary 3.3** **UNDIRECTED ST-REALIZABILITY** with no  $\epsilon$  edges is **SLogCFL**-complete.

## B Proofs

**Theorem 2.2** **ST-REALIZABILITY** is **LogCFL**-complete.

*Proof.* We first show that ST-REALIZABILITY is in **LogCFL**. Let  $\langle \mathcal{G}(V, E), s, t \rangle$  be an instance of ST-REALIZABILITY. An AuxPDA (say  $\mathcal{M}$ ) deciding ST-REALIZABILITY operates by starting at node  $s$  and nondeterministically guessing the nodes of a directed path from  $s$  to  $t$ .  $\mathcal{M}$  records the position of the current node at each step on the work tape. If the current node is  $u$ ,  $\mathcal{M}$  nondeterministically selects the next node  $v$  such that  $(u, v)$  is a directed edge in  $H$ . Let the labels of  $u$  and  $v$  be  $\alpha_u$  and  $\alpha_v$  respectively. If  $(u, v)$  is labeled *push*,  $\mathcal{M}$  pushes  $\alpha_v$  onto its pushdown tape. If  $(u, v)$  is labeled *pop*, it pops  $\alpha_u$  from its pushdown tape and verifies that the new symbol on the stack is  $\alpha_v$ . If not, it terminates and *rejects*. If  $(u, v)$  is labeled  $\epsilon$  then  $\mathcal{M}$  checks if  $\alpha_u$  and  $\alpha_v$  are equal. If not, it terminates and *rejects*.  $\mathcal{M}$  repeats this action until it reaches node  $t$  with an empty pushdown tape and *accepts*, or until it has gone on for  $N$  steps and *rejects*, where  $|V| = N$  is the number of nodes in  $\mathcal{G}$ . Hence ST-REALIZABILITY is in **LogCFL**.

We now show a log space reduction from any language  $\mathcal{L}$  in **LogCFL** to ST-REALIZABILITY. Let  $\mathcal{M}$  be an AuxPDA deciding  $\mathcal{L}$  in log space. Given an input  $w$ , we construct a directed graph  $\mathcal{H}$  along with the vertex and labels and two special vertices  $s$  and  $t$  such that  $\mathcal{H}$  has a *realizable* path from  $s$  to  $t$  if and only if  $\mathcal{M}$  accepts  $w$ .

The nodes of  $\mathcal{H}$  are the configurations of  $\mathcal{M}$  on  $w$ . For configuration  $A$  and  $B$  of  $\mathcal{M}$  on  $w$ , the pair  $(A, B)$  is an edge of  $\mathcal{H}$  if  $B$  is one of the possible next configurations of  $\mathcal{M}$  starting at  $A$ . We say that  $A$  *yields*  $B$ . The vertices of  $\mathcal{H}$  are labeled with the topmost symbol of the stack in the corresponding configuration of  $\mathcal{M}$ . The edge  $(A, B)$  is labeled *push* (resp. *pop*) if  $\mathcal{M}$  performs a *push* (resp. *pop*) operation to reach from  $A$  to  $B$ . If  $\mathcal{M}$  reaches from  $A$  to  $B$  without a push or pop then the edge  $(A, B)$  is labeled  $\epsilon$ . Node  $s$  is the start configuration of  $\mathcal{M}$  on  $w$ . We may assume that  $\mathcal{M}$  has a unique accepting configuration, and we designate this configuration to be node  $t$ . This mapping reduces  $\mathcal{L}$  to ST-REALIZABILITY because, whenever  $\mathcal{M}$  accepts  $w$ , some branch of its computation accepts, which corresponds to a realizable path from  $s$  to  $t$  in  $\mathcal{H}$ . Conversely, if some realizable path exists from  $s$  to  $t$  in  $\mathcal{H}$ , some computation branch accepts when  $\mathcal{M}$  runs on input  $w$ . The reduction can be performed by a log space transducer which, on input  $w$ , outputs a description of  $\mathcal{H}$  along with the vertex and edge labels.  $\square$

Corollary 2.3 ST-REALIZABILITY with no  $\epsilon$  edges is **LogCFL**-complete.

*Proof.* We replace each directed edge  $(u, v)$  labeled with  $\epsilon$  with two directed edges  $(u, w)$  and  $(w, v)$ , where  $w$  is a new node. The label of  $(u, w)$  (resp.  $(w, v)$ ) is set to *push* (resp. *pop*). Repeat this for every edge, adding a new node every time. We introduce a new label  $\alpha_{k+1}$  and label all the new nodes with  $\alpha_{k+1}$ . It is easy to see that a path from  $s$  to  $t$  is realizable in the original graph if and only if it is realizable in the new graph. Hence ST-REALIZABILITY reduces to ST-REALIZABILITY with no  $\epsilon$  edges.

Equivalently, we may assume that an AuxPDA always pushes or pops a symbol at every step. If an AuxPDA doesn't push or a pop at every step then we introduce an extra alphabet in its stack alphabet which is pushed onto the stack when nothing is done to the stack. This alphabet is first popped before performing a valid stack move.  $\square$

Theorem 4.4 **NL = 1LogCFL**.

*Proof.* **1LogCFL**  $\subseteq$  **NL**: An **NL**-machine (say  $\mathcal{M}$ ) non-deterministically guesses an  $s$ - $t$  path (say  $P$ ).  $\mathcal{M}$  traverses the edges along  $P$  and maintains a counter  $C$ .  $\mathcal{M}$  increments (resp. decrements)  $C$  if the current edge is labeled *push* (resp. *pop*). If  $C$  was ever negative then  $\mathcal{M}$  rejects.  $\mathcal{M}$  accepts iff  $C = 0$  when it reaches  $t$ .

**NL**  $\subseteq$  **1LogCFL**: We replace each directed edge (say  $(u, v)$ ) of ST-CONNECTIVITY by two directed edges  $(u, w)$  and  $(w, v)$  and label them *push* and *pop* respectively. We add a new vertex  $w$  for each edge  $(u, v)$ . There is an  $s$ - $t$  path in the original graph iff there is a realizable path (according to the definition from Section 4.2.1) in the modified graph.  $\square$

Theorem 4.6 **BALANCED ST-CONNECTIVITY** is **1SGSLogCFL**-complete.

*Proof.* **BALANCED ST-CONNECTIVITY**  $\in$  **1SGSLogCFL**: Let  $\mathcal{G}(V, E)$  be an instance of **BALANCED ST-CONNECTIVITY**. Let  $\mathcal{G}'(V, E')$  be the underlying undirected graph of  $\mathcal{G}$ . If  $(u, v) \in E$  and  $(v, u) \in E$  then we label the edges  $(u, v)$  and  $(v, u)$  of  $\mathcal{G}'$  with  $\epsilon$ . If  $(u, v) \in E$  and  $(v, u) \notin E$  then we label the edge  $(u, v)$  of  $\mathcal{G}'$  with *push* and label the edge  $(v, u)$  of  $\mathcal{G}'$  with *pop*. Note that the edge labels of  $\mathcal{G}'$  are symmetric. There is a balanced  $s$ - $t$  path in  $\mathcal{G}$  iff there is a realizable  $s$ - $t$  path (according to the definition from Section 4.2.3) in  $\mathcal{G}'$ .

**BALANCED ST-CONNECTIVITY** is **1SGSLogCFL**-hard: An instance of **1SGSLogCFL** is an undirected graph (say  $G$ ) with edges labeled from  $\{\textit{push}, \textit{pop}, \epsilon\}$ . These edge labels are symmetric as defined in Section 3. We construct a directed graph  $H$  on the same vertex set. If the edge  $(u, v)$  of  $G$  is labeled  $\epsilon$  we add the edges  $(u, v)$  and  $(v, u)$  in  $H$ . If the edge  $(u, v)$  is labeled *push* (by symmetry the edge  $(v, u)$  is labeled *pop*) we add a directed edge  $u, v$  in  $H$ . There is a realizable  $s$ - $t$  path in  $G$  iff there is a balanced  $s$ - $t$  path in  $H$ .  $\square$

Theorem 4.7 **POSITIVE BALANCED ST-CONNECTIVITY** is **1SLogCFL**-complete.

*Proof.* Similar to the proof of Theorem 4.6.  $\square$

Theorem 5.2 Let  $\mathcal{G}$  be an instance of ST-REALIZABILITY.  $\mathcal{G}^* = \langle \Upsilon^*, \mathcal{E}^* \rangle$  can be computed using  $O(\log n)$  repeated applications of **Square**( $\mathcal{G}$ ).

*Proof.* We first state the relevant definitions and lemmas from [NR95]. A *path description* is a triple  $(A, B, i)$  consisting of two surface configurations  $A$  and  $B$  and an even natural number  $i$ . A description is *realizable* if  $A$  and  $B$  are realizable. By Corollary 2.3 we may assume that there are no  $\epsilon$  edges in an instance of ST-REALIZABILITY, and hence  $i$  can only be an even number. In particular,  $(A, B, i)$  represents several paths of length  $i$  between  $A$  and  $B$ .

The relation  $\vdash$  shows how to split computation paths recursively into shorter and shorter paths until we end up with trivial paths. Let  $x = (A, B, i)$ ,  $y = (C, D, j)$ , and  $z = (E, B, k)$  be path descriptions. Then we write  $y, z \vdash x$  and  $z, y \vdash x$  if and only if

- (1) the level of the pushdown is equal for  $A, E$  and  $B$ ;
- (2) there exists a computation from  $A$  to  $C$  in one step, pushing a symbol  $\alpha$  onto the pushdown tape during this step;
- (3) there exists a computation from  $D$  to  $E$  in one step, popping  $\alpha$  from the pushdown tape; and
- (4)  $j + k = i - 2$ .

Note that identical pushdown heights of  $A, E$  and  $B$  imply that  $C$  and  $D$  have same pushdown height. Also,  $j$  and  $k$  are always even. In this way we can reduce the checking of realizability of  $x$  to the checking of the realizability of smaller paths  $y$  and  $z$ . We now state two crucial lemmas from [NR95] that gives a

“balanced” partition of realizable computation. The proofs of these lemmas are based on a *recursive descent* using the properties of the decomposition relation  $\vdash$ .

**Lemma B.1.** (Niedermeier and Rossmannith [NR95]) Let  $(A, B, i)$  denote a realizable path description for a fixed computation path of length  $i \geq 2$  between  $A$  and  $B$ . Then there exist uniquely determined subpaths  $(C, D, i_1)$ ,  $(E, F, i_2)$  and  $(G, D, i_3)$  of  $(A, B, i)$  such that  $(E, F, i_2), (G, D, i_3) \vdash (C, D, i_1)$  and  $i_2, i_3 \leq i/2 < i_1$ .

Lemma B.1 splits a fixed computation path into three paths. The first two paths are the subpaths  $(E, F, i_2)$  and  $(G, D, i_3)$  and the third one is the path  $(A, B, i)$  with gap  $(C, D, i_1)$ . This means that the verification of the realizability of  $(A, B, i)$  can be reduced to showing that  $(E, F, i_2)$ ,  $(G, D, i_3)$  and the pair-with-gap  $(A, (C, D, i_1), B, i)$  are realizable.

A description for a *path with gap*  $(A, (C, D, j), B, i)$  consists of four surface configurations  $A, B, C, D$  and two even numbers  $i$  and  $j$  with  $j \leq i$ . A path with gap  $(A, (C, D, j), B, i)$  is called *realizable* iff  $(A \rightsquigarrow (C, D) \rightsquigarrow B)$  and there exists a computation path from  $A$  to  $C$  and one from  $D$  to  $B$  with total number of steps  $j - i$ . Now we generalize the decomposition relation  $\vdash$  to computation paths with gap. Let  $x = (A, (C, D, j), B, i)$  and, first, let  $y = (E, (C, D, j), F, k)$  and  $z = (G, B, l)$  or, second, let  $y = (E, F, k)$ ,  $z = (G, (C, D, j), B, l)$ . Then we write  $y, z \vdash x$  and  $z, y \vdash x$  if and only if

- (1) the level of the pushdown is equal for  $A, G$  and  $B$ ;
- (2) there exists one step from  $A$  to  $E$  pushing a symbol  $\alpha$  onto the pushdown tape;
- (3) there is one step from  $F$  to  $G$  popping  $\alpha$  from the pushdown tape; and
- (4)  $k + l = i - 2$ .

The following lemma is the analogue to Lemma B.1 for a fixed computation path with gap.

**Lemma B.2.** (Niedermeier and Rossmannith [NR95]) Let  $(A, (C, D, j), B, i)$ ,  $i - j \geq 2$  denote a realizable path with gap. Then there exist uniquely determined paths  $y = (E, (C, D, j), F, i_1)$  and either

(1)  $z_1 = (G, (C, D, j), H, i_2)$  and  $z_2 = (I, F, i_3)$ , such that  $z_1, z_2 \vdash y$  and  $i_2 - j \leq (i - j)/2 < i_1 - j$  or

(2)  $z_1 = (G, H, i_2)$  and  $z_2 = (I, (C, D, j), F, i_3)$ , such that  $z_1, z_2 \vdash y$  and  $i_3 - j \leq (i - j)/2 < i_1 - j$ .

Lemma B.2 is used to decompose *paths with gaps* in a balanced way. To check the realizability of  $(A, (C, D, j), B, i)$  we examine the realizability of  $(A, (E, F, i_1), B, i)$ ,  $z_1$  and  $z_2$ . Both possible subpaths with gap have length less than or equal to half of the length of the whole path with gap  $(A, (C, D, j), B, i)$ . The arising subpath without gap may have a maximum length of  $i - j - 2$  and will be split in a balanced way using Lemma B.1.

---

**Square** $(\langle \Upsilon, \mathcal{E} \rangle)$

for all  $a, b \in V$  update  $\mathcal{E}$  as follows :

$$\mathcal{E}[a, b] = \sum_{c, e, f, g, d} \Upsilon[a, (c, d), b] \cdot \Upsilon[c, (e, f), g] \cdot \mathcal{E}[e, f] \cdot \mathcal{E}[g, d]$$

for all  $a, b, c, d \in V$  update  $\Upsilon$  as follows :

$$\begin{aligned} \Upsilon[a, (c, d), b] &= \sum_{c', e', f', g', d'} \Upsilon[a, (c' d'), b] \cdot \Upsilon[c', (e', f'), g'] \cdot \Upsilon[e', (c, d), f'] \cdot \mathcal{E}[g', d'] \\ &+ \sum_{c', e', f', g', d'} \Upsilon[a, (c' d'), b] \cdot \Upsilon[c', (e', f'), g'] \cdot \mathcal{E}[e', f'] \cdot \Upsilon[g', (c, d), d'] \end{aligned}$$

**return**  $\langle \Upsilon, \mathcal{E} \rangle$

---

**Square** $\langle \Upsilon, \mathcal{E} \rangle$

$$\mathcal{E} = (\Upsilon \otimes_2 ((\Upsilon \otimes_2 \mathcal{E}) \otimes_1 \mathcal{E}))$$

$$\Upsilon = (\Upsilon \otimes_5 ((\Upsilon \otimes_5 \Upsilon) \otimes_3 \mathcal{E})) + (\Upsilon \otimes_5 ((\Upsilon \otimes_2 \mathcal{E}) \otimes_4 \Upsilon))$$

**return**  $\langle \Upsilon, \mathcal{E} \rangle$

---

Our **Square** algorithm is based on Lemma B.1 and Lemma B.2. Since the summation is taken over all possible intermediate surface configurations, the matrices  $\mathcal{E}$  and  $\Upsilon$  are populated in a *bottom-up* manner. Based on the above lemmas, Niedermeier and Rossmannith [NR95] constructed an **SAC**<sup>1</sup> circuit simulating the corresponding AuxPDA. The circuit consists of gates denoted by  $\langle A, B, i \rangle$  and  $\langle A, (C, D, j), B, i \rangle$  that compute the realizability of the corresponding path descriptions. Our **Square** algorithm is inspired by their approach. Translating the sum symbols into (unbounded) OR-gates and multiplication symbols into (bounded) AND-gates we get the corresponding **SAC**<sup>1</sup> circuit for a given vertices  $s$  and  $t$  of the graph  $\mathcal{G}$ . Each **Square** operation on the graph  $\mathcal{G} = \langle \Upsilon, \mathcal{E} \rangle$  reduces the depth of the corresponding circuit by  $O(1)$ . Since our squaring operation is used to update all the entries of  $\mathcal{E}$  and  $\Upsilon$ , after  $O(\log n)$  repeated squaring operations, we can decide the  $s$ - $t$  realizability for any two given vertices  $s$  and  $t$ . Similar argument holds for paths with gap.

Hence, we can compute the transitive closure  $\mathcal{G}^* = \langle \Upsilon^*, \mathcal{E}^* \rangle$  using  $O(\log n)$  repeated squaring operations on  $\mathcal{G} = \langle \Upsilon, \mathcal{E} \rangle$ .  $\square$

**Theorem 5.3** Let  $\mathcal{G}$  be an instance of ST-REALIZABILITY.  $\mathcal{G}^* = \langle \Upsilon^*, \mathcal{E}^* \rangle$  can be computed using  $O(\log n)$  repeated applications of **SimpleSquare** $(\mathcal{G})$ .

*Proof.* For realizable paths (both standard and gap paths) of length at most four, it is easy to verify that an application of **SimpleSquare** reduces the path length by a factor of at least  $\frac{3}{4}$ . For paths of length greater than four, we divide the path into three smaller paths using Lemma B.1 for standard paths and Lemma B.2 for path with gaps and use induction. This implies that one application of **SimpleSquare** reduces the path length by a constant factor. Hence  $O(\log n)$  repeated applications of **SimpleSquare** $(\mathcal{G})$  suffice to compute the transitive closure  $\mathcal{G}^*$ .  $\square$

**Theorem 6.4** The algorithm **Connect** finds  $\mathcal{G}^* = \langle \Upsilon^*, \mathcal{E}^* \rangle$  in parallel time  $O(\log^2 n)$  using  $n^4$  processors in the CREW PRAM model.

*Proof.* The following observations state that the hooking process creates pseudotrees on vertices from  $V$  and  $V^2$ . A closer look at the contraction step reveals that it is essentially simulating **SymmetricSquare**. Hence the overall running time of the contraction step is  $O(\log n)$  time.

**Observation :** Let  $V_s \subseteq V$  denote an  $s$ -component of  $\mathcal{G}$  such that  $|V_s| \geq 2$  and define the function  $C : V_s \rightarrow V_s$  by  $C(i) = \mathbf{StandardHook}(i)$ . The function  $C$  defines a directed graph  $G_s(C) = (V_s, F)$  where  $F = \{(i, C(i)) \mid i \in V_s\}$ . Then  $G_s(C)$  is a collection of pseudotrees with circumference one, and the smallest-numbered vertex in each pseudotree is in the cycle of the pseudotree.

**Observation :** Let  $V_g \subseteq V^2$  denote a  $g$ -component of  $\mathcal{G}$  such that  $|V_g| \geq 2$  and define the function  $C : V_g \rightarrow V_g$  by  $C(i, j) = \mathbf{GapHook}(i, j)$ . The function  $C$  defines a directed graph  $G_g(C) = (V_g, F)$  where  $F = \{((i, j), C(i, j)) \mid (i, j) \in V_g\}$ . Then  $G_g(C)$  is a collection of pseudotrees with circumference one, and the lexicographically smallest vertex in each pseudotree is in the cycle of the pseudotree.

**Time and Processor Bounds :** The main loop of the **Connect** program is executed  $O(\log n)$  times. Within the loop, the iteration at step 14 is executed  $O(\log n)$  times. Thus the algorithm requires  $\Omega(\log^2 n)$  time. Steps 3, 12, 18 require  $O(1)$  time using  $\Omega(n)$  processors. Steps 4, 13, 19 require  $O(1)$  time using  $\Omega(n^2)$  processors. Steps 14-17 require  $O(\log n)$  time using  $\Omega(n^2)$  processors. **StandardHook** and **GapHook** are essentially computing minimum of at most  $O(n^2)$  integers (accessing both  $\mathcal{E}$  and  $\Upsilon$ ) and hence can be programmed to execute in  $O(\log n)$  time using  $O(n^2)$  processors. Hence the total running time is  $O(\log^2 n)$ . The total number of processors used is  $O(n^4)$ . The correctness of the algorithm follows from Corollary 6.3.  $\square$

## C BALANCED ST-CONNECTIVITY

In all the algorithms presented in this paper we are only looking for balanced paths of length at most  $n$ . The example in Figure 2 shows an instance of **BALANCED ST-CONNECTIVITY** where the *only* balanced path between  $s$  and  $t$  is of length  $\Theta(n^2)$ . The directed simple path from  $s$  to  $t$  is of length  $n/2$ . There is a cycle of length  $n/2$  at the vertex  $v$ . All the edges (except  $(v, u)$ ) on this cycle are undirected. The balanced path from  $s$  to  $t$  is obtained by traversing from  $s$  to  $v$ , traversing the cycle clockwise for  $n/2$  times and then traversing from  $v$  to  $t$ . This path is not simple.

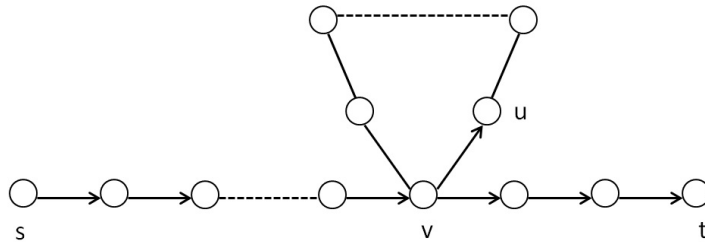


Figure 2: Example.

We now define two more connectivity problems. A path  $P$  from  $s$  to  $t$  is called  $k$ -balanced if the number

of forward edges along  $P$  minus the number of backward edges along  $P$  is equal to  $k$ . A path  $P$  from  $s$  to  $t$  is called *positive  $k$ -balanced* if  $P$  is *positive* and  *$k$ -balanced*.

**K-BALANCED ST-CONNECTIVITY** : Given a directed graph  $\mathcal{G}(V, E)$  and two distinguished nodes  $s$  and  $t$ , decide if there is  *$k$ -balanced* path (of length at most  $n$ ) between  $s$  and  $t$ .

**POSITIVE K-BALANCED ST-CONNECTIVITY** : Given a directed graph  $\mathcal{G}(V, E)$  and two distinguished nodes  $s$  and  $t$ , decide if there is *positive  $k$ -balanced* path (of length at most  $n$ ) between  $s$  and  $t$ .

K-BALANCED ST-CONNECTIVITY can be solved as follows : Add a new vertex  $t'$  and a directed path (with all new vertices) of length  $k$  from  $t'$  to  $t$ . Find a balanced path from  $s$  to  $t'$  in this modified graph. POSITIVE K-BALANCED ST-CONNECTIVITY can be solved similarly.