Average Case Complexity of Unbounded Fanin Circuits

Andreas Jakoby[1,2] and Rüdiger Reischuk[2]
[1] Department of Computer Science, University of Toronto
[2] Institut für Theoretische Informatik, Med. Universität zu Lübeck

January 2000

Abstract

Several authors have shown that the PARITY-function cannot be computed by unbounded fanin circuits of small depth and polynomial size. Even more, constant depth k circuits of size $\exp(n^{\Theta(1/k)})$ give wrong results for PARITY for almost half of all inputs.

We generalize these results in two directions. First, we obtain similar tight lower bounds for the average case complexity of circuits, measuring the computational delay instead of the static circuit depth.

Secondly, with respect to average delay of unbounded fanin circuits we completely classify all parallel prefix functions, for which PARITY is just one promiment example. It is shown that only two cases can occur: a parallel prefix functions f either has the same average complexity as PARITY, that is the average delay has to be of order $\Theta(\log n/\operatorname{loglog} s)$ for circuits of size s, or f can be computed with *constant* average delay and almost linear size – there is no complexity level in between.

This classification is achieved by analyzing the algebraic structure of semigroups that correspond to parallel prefix functions. It extends methods developed for bounded fanin circuits by the first author in his Ph.D. Thesis.

1 Introduction

To analyse the average-case complexity of Boolean functions a dynamic complexity measure for Boolean circuits based on the notion of delay has been introduced in [JRS94]. For the classical circuit model with OR, AND and NOT gates of bounded fanin and fanout we have shown that n-ary functions like the OR, the addition or threshold functions with a fixed threshold can be computed much faster on the average compared to the worst-case for which the trivial $\log n$ depth lower bound holds. Other functions like PARITY do not allow a significant speedup of the average time complexity, regardless of the size of the circuits. In [JRSW94] we have generalized our investigations to computing arbitrary prefix functions. Tight bounds for the average complexity have been obtained by analysing the algebraic structure of the semigroups that correspond to the prefix operator. These results have been extended in [Jak98, Jak99] to parallel prefix computations in other models like linear arrays and networks of automata.

A different line of research has considered the complexity of Boolean functions in a circuit model where gates of arbitrary large fanin can be used. It is well known that in this model every Boolean function can be computed in constant depth by circuits of exponential size. Functions like the OR or the addition again can be computed much more efficient, that is in constant depth and polynomial size [ChFL83a]. But PARITY remains difficult – the best speedup for polynomial size one can obtain is depth $\Omega(\log n/\operatorname{loglog} n)$ [Yao85]. Håstad has proven optimal lower size bounds with respect to the circuit depth by applying the probabilistic restriction method very precisely. Babai and Cai have shown that circuits of constant depth k and size $\exp(n^{\Theta(1/k)})$ generate wrong results for PARITY for almost half of all inputs [Bab87, Cai89]. Smolensky has presented an algebraic approach to generalize the PARITY lower bound to arbitrary $\bmod\ m$-functions.

In this paper we combine both questions and investigate whether PARITY can be computed more efficiently by unbounded fanin circuits on the average. That means, instead of the static complexity measure depth we consider the dynamic measure delay. Krapchenko has given an example showing that an increase of the circuit size cannot be avoided in general when a delay minimal circuit is transformed into a circuit of minimal depth [Kra78]. Thus, lower bounds on circuit depth, even the ones in [Bab87, Cai89] for almost half of all inputs, do not directly translate into bounds for the average delay. At the same time we try to classify Boolean functions according to their average case complexity in the unbounded fanin model. Our work starts by investigating algebraic properties of the corresponding semigroups in a similar spirit as in [JRSW94, Jak98] now tailored to unbounded fanin computations.

For the worst-case complexity Chandra, Fortune and Lipton have investigated the algebraic structure of prefix operators and shown that constant depth and polynomial size cannot be achieved simultaneously if the semigroup contains a subgroup [ChFL83a]. These results cannot simply be translated to the average case because they do not hold under this setting. Consider the following simple example of an extension of the PARITY operator $\oplus$, which still contains the 2-element group as a subgroup. Let us call this function *parity with reset*, $\otimes : \{0, 1, \mathtt{r}\}^* \to \{0, 1\}^*$, defined by

$$\otimes(w) \quad := \quad \begin{cases} 0 & \text{if } w = \lambda \text{ or } w[|w|] = \mathtt{r} \ , \\ \oplus(\otimes(w[1; |w| - 1]), w[|w|]) & \text{else.} \end{cases}$$

An input value $\mathtt{r}$ resets the value of the current prefix to 0. Since a circuit that computes the *parity with reset* function can also be used to compute PARITY, the function $\otimes$ cannot be computed by a polynomial size circuit with depth $o(\log n / \operatorname{loglog} n)$. On the other hand, our results imply that $\otimes$ can be computed by a circuit of nearly linear size within a constant average delay.

Similarly, the classification of Bilardi, Preparata for constant depth and linear size prefix circuits based on the notion of monoidel cycles [BiPr90] does not extend to the average case.

The second major step in our analysis of the problem is a nontrivial extension of the probabilistic restriction method to arbitrary finite domains. This way, we can deduce bounds on the error probabilities of average-case delay bounded circuits with unbounded fanin gates, from which the lower bounds for prefix functions of certain types will follow.

The results are as follows. For parity and other prefix functions with a group like structure "embedded" in the semigroup (in a sense to be made precise below) the same lower bounds are obtained for the average time complexity as in the worst case. For all other functions we construct polynomial size circuits with constant average delay. Thus, we obtain a strict separation with a large complexity gap from constant to $\Theta(\log n / \operatorname{loglog} n)$.

The paper is organized as follows. After introducing the necessary formal models and notation we state the main results precisely. Then, we outline the essential techniques and arguments to obtain the upper and lower bounds. Some of the technical proofs are postponed to the appendix.

2 Definitions

Let $\Sigma, \Sigma_I, \Sigma_O$ be finite alphabets, and let $\boldsymbol{\lambda}$ denote the empty word. In the following, Σ_I and Σ_O will typically be subsets of Σ. For a string $\boldsymbol{w = w[1]w[2] \dots w[n]} \in \Sigma^*$ of length $|w| = n$ and for $1 \leq i \leq j \leq n$ define $\boldsymbol{w[i;j]} := w[i] \dots w[j]$. To shorten notion we set $\mathbf{llog}\, \boldsymbol{n} := \operatorname{loglog} n$.

2.1 Circuit Delay

The depth of a circuit is a static and thus worst-case measure for the (parallel) time complexity of computing functions in the circuit model. In [JRS94] we have given a definition for the delay of circuits and used this notion to define an average-case measure for the time complexity in the circuit model.

Definition 1 *Let C be a circuit over the standard basis $\{\mathtt{OR}, \mathtt{AND}, \mathtt{NOT}\}$, where the fanin of* OR- *and* AND-*gates may be bounded or unbounded, and let $res_v(x)$ denote the value computed by gate v on input x. For input gates and constant gates v set $\mathrm{time}_v(x) := 0$.*
For an internal nonconstant gate v with p direct predecessors $v_1, \dots, v_p$ we define

$$\mathbf{time}_{\boldsymbol{v}}\boldsymbol{(x)} \quad := \quad 1 \ + \ \min\{t \mid \textit{the values } res_{v_i}(x) \textit{ with } \mathrm{time}_{v_i}(x) \leq t \textit{ uniquely determine } res_v(x)\} \ .$$

For the circuit C itself with output gates $y_1, \dots, y_m$ the global time function is given by

$$\mathbf{time}_{\boldsymbol{C}}\boldsymbol{(x)} \quad := \quad \max_i \ \mathrm{time}_{y_i}(x) \ .$$

In order to compute functions over arbitrary finite domains Σ_I in the Boolean circuit model, symbols $\sigma \in \Sigma_I$ will be coded in an appropriate way by binary vectors. A circuit C is said to compute a function $f : \Sigma_I^n \to \Sigma_O^m$ if C maps the corresponding code vectors. For a function $f : \Sigma_I^* \to \Sigma_O^*$ let f_n denote the restriction of f to Σ_I^n and let $\mathbf{UbCir}(\boldsymbol{f_n, s})$ denote the set of all circuits with unbounded fanin and fanout over the standard basis (**ub-circuits** for short) of size at most s that compute f_n. As usual, f will be computed by a family of circuits $C = C_1, C_2, \dots$ with C_n responsible for f_n. We will analyse the average case complexity of functions over Σ_I^* with respect to the family of **uniform probability distributions** $(\mu_n^{\mathrm{uni}})_{n=1,2,\dots}$, where μ_n^{uni} assigns probability $|\Sigma_I|^{-n}$ to every string over Σ_I of length n.

Definition 2 *For a circuit C_n with inputs from Σ_I^n define the* **expected time**, *resp. the* **worst-case time** *as*

$$\mathbf{etime}(\boldsymbol{C_n}) \;:=\; \sum_{x\in\Sigma_I^n} \mu_n^{\text{uni}}(x)\cdot \text{time}_{C_n}(x) \quad \textit{and} \quad \mathbf{time}(\boldsymbol{C_n}) \;:=\; \max_{x\in\Sigma_I^n} \text{time}_{C_n}(x)\ .$$

2.2 Prefix Functions

Definition 3 *A* **groupoid** $\langle\Sigma,\otimes\rangle$ *consists of a set Σ (in the following also called alphabet) and a binary operator $\otimes:\ \Sigma\times\Sigma\to\Sigma$, which does not have to be associative.*
A binary operator will be extended to strings of arbitrary length ℓ in the natural way: for $\ell > 2$ and $w\in\Sigma^\ell$

$$\otimes\boldsymbol{(w)} \;:=\; (\ldots((w[1]\otimes w[2])\otimes w[3])\otimes\ldots w[\ell])\ ,$$

and for $|w|\le 1$ $\otimes(w):=w$. Often the first symbol of a word over Σ will have special importance. Thus for $q\in\Sigma\cup\{\lambda\}$ and $w\in\Sigma^\ell$, when considering the word $q\,w$ we will also use the notation $\otimes_{\boldsymbol{q}}\boldsymbol{(w)} := \otimes(qw)$. Extending this notion to sets $Q\subseteq\Sigma$ and $X\subseteq\Sigma^*$ we define $\otimes(Q,X) := \{\otimes_q(w)\mid q\in Q,\ w\in X\}$.

Definition 4 *For a groupoid $\langle\Sigma,\otimes\rangle$ and a subset $\Sigma_I\subseteq\Sigma$ we define the n-ary* **parallel prefix function** $\mathbf{PP}^{\boldsymbol{n}}_{\langle\boldsymbol{\Sigma},\otimes\rangle,\boldsymbol{\Sigma_I}}:\ \Sigma_I^n\to\Sigma^n$ *as follows. An input $x=x[1]\ldots x[n]$ is transformed into the vector $y=y[1]\ldots y[n]$ with*

$$y[i] := \otimes(x[1;i])\ .$$

If the domain Σ_I equals Σ we simply write $\mathtt{PP}^n_{\langle\Sigma,\otimes\rangle}$. A function $f:\ \Sigma_I^*\to\Sigma_O^*$ is called a **prefix function** *if there exists a groupoid $\langle\Sigma,\otimes\rangle$ such that f restricted to Σ_I^n is identical to $\mathtt{PP}^n_{\langle\Sigma,\otimes\rangle,\Sigma_I}$. Note that this implicitly requires that $|f(x)|=|x|$ for all arguments x of f and that Σ contains both Σ_I and Σ_O. The set of all prefix functions will be denoted by $\boldsymbol{\mathcal{F}}_{\mathbf{prefix}}$.*
A groupoid $\langle\Sigma,\otimes\rangle$ is called **minimal** *for a prefix function f if for any pair $q_1,q_2\in\Sigma$ there exists a string $w\in\Sigma_I^*$ such that $\otimes_{q_1}(w)\neq\otimes_{q_2}(w)$. Minimal groupoids are unique up to isomorphisms. In the following, for prefix functions f we will only consider minimal groupoids, which will be denoted by $\langle\boldsymbol{\Sigma^{[f]}},\boldsymbol{\otimes^{[f]}}\rangle$.*

Groupoids can be represented by directed graphs which somehow resemble the transition diagrams of finite automata. Symbols in Σ correspond to the states of the machine.

Definition 5 *For a groupoid $\langle\Sigma,\otimes\rangle$ and an input alphabet $\Sigma_I\subseteq\Sigma$ define the* **transition graph** $\boldsymbol{G_{\langle\Sigma,\otimes\rangle,\Sigma_I}} = (\Sigma,E)$ *by*

$$\forall q_1,q_2\in\Sigma\ :\quad (q_1,q_2)\in E \quad\Longleftrightarrow\quad \exists\ q_3\in\Sigma_I\ :\ q_1\otimes q_3 = q_2\ .$$

If this graph is strongly connected we call the groupoid **closed**.
For a prefix function f let $\boldsymbol{G_f}$ denote the transition graph of the minimal groupoid $\langle\Sigma,\otimes\rangle$ for f. If G_f is closed we will also call f closed. The set of all closed prefix functions will be denoted by $\boldsymbol{\mathcal{F}}^{\mathbf{closed}}_{\mathbf{prefix}}$.

A concept similar to transition graphs of groupoids, but requiring a distinguished starting state has been introduced under the name Cayley graphs (see for example [BiPr89]).

2.3 Confluences and Diffluences

Based on the minimal groupoid $\langle\Sigma,\otimes\rangle$ of a prefix function f and its transition graph G_f we will now analyse some basic properties, which turn out to be helpful for analysing the average-case complexity of f. For $t\in\mathbb{N}$, the set of nodes of G_f that can be reached in exactly t steps starting at node $q\in\Sigma$ will be denoted by $\boldsymbol{R_{f,q}(t)} := \otimes(q,\Sigma_I^t)$. Furthermore, for $t\ge 1$ let $\boldsymbol{R_f(t)} := \otimes(\lambda,\Sigma_I^t)$. It is easy to see that these **reachability sets** share the following property:

$$\forall\ t_1,t_2\in\mathbb{N}\ :\qquad R_{f,q}(t_1)=R_{f,q}(t_2) \qquad\Longleftrightarrow\qquad R_{f,q}(t_1+i)=R_{f,q}(t_2+i)\ \forall i\in\mathbb{N}\ .$$

Since there are at most $2^{|\Sigma|}$ different possibilities for $R_{f,q}(t)$ this relation implies that there exist numbers $\tau,\pi\in\{1,2,\ldots,2^{|\Sigma|}\}$ with $R_f(\tau)=R_f(\tau+i\cdot\pi)$ for all $i\in\mathbb{N}$. Let us call the smallest such τ the **start of periodicity, $\boldsymbol{\tau(f)}$**, and the smallest such π the **period, $\boldsymbol{\pi(f)}$**, of f. In [Jak98] a significantly better bound than $\tau(f)\le 2^{|\Sigma|}$ has been shown.

Proposition 1 *For every prefix function $f\in\mathcal{F}_{\text{prefix}}$ the start of periodicity can be bounded by $\tau(f)\le|\Sigma|^2+|\Sigma|$. For $f\in\mathcal{F}^{\text{closed}}_{\text{prefix}}$ it holds in addition: If we augment the parallel prefix function by a symbol $q\in\Sigma$ at the beginning, that is if we consider the function $\mathtt{PP}_{\langle\Sigma,\otimes_q\rangle,\Sigma_I}$, then the periodicity $\pi(f)$ is independent of q and can be bounded by $|\Sigma|$.*

The periodic or nonperiodic behavior of the transition graph turns out to be an important classification tool.

Definition 6 *For $q_1, q_2 \in \Sigma$, a string $w \in \Sigma_I^*$ with $\otimes_{q_1}(w) = \otimes_{q_2}(w)$ is called a* **confluence**. *$w \in \Sigma_I^*$ is a confluence for a subset $\Sigma' \subseteq \Sigma$ if $|\otimes(\Sigma', w)| = 1$. f will be called* **confluent** *if $R_f(\tau(f))$ has a confluence. Let $\boldsymbol{\alpha(f)}$ denote the minimal length of such a confluence w. Let $\boldsymbol{\mathcal{F}_{\mathrm{conf}}}$ denote the set of all confluent functions in $\mathcal{F}_{\mathrm{prefix}}$.*
A prefix function f is called **strictly confluent** *if there exists some $\ell \in \mathbb{N}$ such that each string $w \in \Sigma_I^\ell$ is a confluence for $R_f(\tau(f))$. Define $\boldsymbol{\mathcal{F}_{\mathrm{str\text{-}conf}}}$ as the set of all strictly confluent functions.*
For a subset $\Sigma' \subseteq \Sigma$, a nonempty string $w \in \Sigma_I^*$ is a **diffluence** *if Σ' contains two distinct elements q_1, q_2 such that $\otimes(\{q_1, q_2\}, w) = \{q_1, q_2\}$. f is* **diffluent** *if $R_f(\tau(f))$ has a diffluence, it is* **strictly diffluent** *if*

$$\exists q_1, q_2 \in R_f(\tau(f))\ \forall w \in \Sigma_I^*\ \exists u \in \Sigma_I^* \ :\ \otimes(\{q_1, q_2\}, wu) \ = \ \{q_1, q_2\}\ .$$

Define $\boldsymbol{\mathcal{F}_{\mathrm{diff}}}$ resp. $\boldsymbol{\mathcal{F}_{\mathrm{str\text{-}diff}}}$ as the set of all (strictly) diffluent functions in $\mathcal{F}_{\mathrm{prefix}}$.

It can be shown that in the definition of strictly confluent the value of ℓ can be bounded by $|\Sigma|^2/2$. In [Jak98, Jak99] is proven that $\mathcal{F}_{\mathrm{str\text{-}conf}}$, $\mathcal{F}_{\mathrm{conf}} \cap \mathcal{F}_{\mathrm{diff}}$, and $\mathcal{F}_{\mathrm{str\text{-}diff}}$ form a partition of $\mathcal{F}_{\mathrm{prefix}}$. Furthermore, we have sown in [JRSW94, Jak98, Jak99] that the confluence property of a prefix function f is strongly related to the average time complexity in bounded degree networks and in the bounded fanin, bounded fanout circuit model. If we investigate circuits with bounded fanin and unbounded fanout we have to extend the notion confluence to the strongly connected components of the transition graph.
A subset $\Sigma' \subseteq \Sigma$ is called a **component** of f if Σ' is a maximal strongly connected component in G_f. For $q \in \Sigma$ let $\boldsymbol{Q[q]}$ denotes the component that contains q. A component Σ' is called **closed** if any node that is reachable from an element of Σ' belongs to Σ'. Let $\boldsymbol{\mathcal{Q}_f^*}$ be the set of all elements of Σ belonging to a closed component of f and $\boldsymbol{\mathcal{Q}_f} := \{\ Q[q] \mid q \in \mathcal{Q}_f^*\ \}$.

3 Main Result

Based on confluences and diffluences $\mathcal{F}_{\mathrm{prefix}}$ can be subdivided into three classes that turn out to be essential for the average case complexity in the bounded fanin circuit model [Jak99].

Theorem A *For any prefix function $f \in \mathcal{F}_{\mathrm{prefix}}$ the expected time for computing f in the bounded fanin and bounded fanout circuit model is either constant iff f is strictly confluent, or of order* $\operatorname{llog} n$ *iff f is confluent and diffluent, or of logarithmic order iff f is strictly diffluent.*

Theorem B *For any prefix function $f \in \mathcal{F}_{\mathrm{prefix}}$ the expected time for computing f in the bounded fanin and unbounded fanout circuit model is either constant iff any closed component of f is strictly confluent, or of order* $\operatorname{llog} n$ *iff every closed component f is confluent and diffluent, or of logarithmic order iff at least one closed component of f is strictly diffluent.*

Considering circuits with unbounded fanin and fanout this classification collapses to only two cases. As the main result of this work we will give an asymptotically tight characterization of prefix functions according to their average time complexity with respect to ub-circuits.

Main Theorem *Let $f \in \mathcal{F}_{\mathrm{prefix}}$ and $s \geq n^{1+\varepsilon}$ for arbitrary $\varepsilon > 0$. Then the average-case complexity of an optimal circuit $C_{n,s} \in \mathrm{UbCir}(f_n, s)$ is given by*

$$\mathrm{etime}(C_{n,s}) \ = \ \begin{cases} \Theta(1), & \textit{if every closed component of } f \textit{ is confluent,} \\ \Theta(\log n/\operatorname{llog} s), & \textit{if there exists a strictly diffluent closed component of } f. \end{cases}$$

For prefix functions like `PARITY` and its generalization for $k > 2$ to sum-mod$k$–functions with $y[i] := 0$ if $\sum_{j=1}^{i} x[j] \equiv 0 \bmod k$, and $y[i] := 1$ else, this implies:

Main Corollary *Polynomial size ub-circuits for $\mathtt{PARITY}_n$ and for n-ary sum- $\bmod\ k$–functions require expected delay $\Theta(\log n/\operatorname{llog} n)$.*

The proof of the lower bound for a strictly diffluent prefix function f is based on a generalization of the probabilistic restriction method. Before going into details we will make a simplification of the set of functions to be investigated. In [Jak99] it has been shown for every prefix function f: if we choose independently and uniformly a string $w = w[1]\ w[2] \ldots w[\ell]$ over Σ_I, where each $w[i]$ has equal probability for the symbols in Σ_I then the probability $p(\ell)$ that $\otimes(w)$ is not an element of a closed component decreases exponentially with ℓ. Hence, by an ub-circuit of constant depth and linear size considering only a prefix of w of logarithmic length we can determine the component $Q[\otimes(w)]$ with high probability. This leads to the following strategy for computing $f(w)$: We partition the input x of length n into a prefix w_1 and a suffix w_2. First, a constant depth linear size ub-circuit C^1 computes the values $f(w_1)$. If the last value, $\otimes(w_1)$, is not an element of a closed component of f we use a standard circuit C^2 of logarithmic depth and linear size to compute $f(x)$. Otherwise, we apply an average efficient ub-circuit C^3 to compute the prefix function $f'(w_2)$ over the groupoid $\langle \Sigma, \otimes_{\otimes(w_1)} \rangle$. The expected delay of the resulting circuit C is bounded by $\mathrm{etime}(C_\ell) \ \le\ \mathrm{depth}(C^1) + p(\ell) \cdot \mathrm{depth}(C^2) + \mathrm{etime}(C^3) \ \le\ \mathrm{etime}(C^3) + O(1)$, if $\ell \le O(\log n)$ is chosen suitably. It thus suffices to analyse the average case complexity of closed prefix functions.

4 Upper Bounds on the Average Delay

In [LaFi80] it is shown that any groupoid can be transformed into a semigroup such that the corresponding prefix functions are equivalent. Hence, it is sufficient to give upper bounds for prefix functions defined over semigroups.

For $\ell \in \mathbb{N}$ define a $\otimes_f^\ell$-gate with $\ell + 1$ inputs $v_0, \ldots, v_\ell$ and outputs $u_0, \ldots, u_\ell$ by

$$\begin{array}{rrcl} & \mathrm{res}_{u_0}(w) & := & (\cdots((\mathrm{res}_{v_1}(w) \otimes \mathrm{res}_{v_2}(w)) \otimes \mathrm{res}_{v_3}(w)) \otimes \cdots) \otimes \mathrm{res}_{v_\ell}(w), \\ \text{and for all } i \in \{1, \ldots, \ell\} & \mathrm{res}_{u_i}(w) & := & (\cdots((\mathrm{res}_{v_0}(w) \otimes \mathrm{res}_{v_1}(w)) \otimes \mathrm{res}_{v_2}(w)) \otimes \cdots) \otimes \mathrm{res}_{v_i}(w) \end{array}$$

To achieve minimal delay, we construct a circuit of $c \cdot n/\ell$ gates of type $\otimes_f^\ell$ as described in Figure 1.

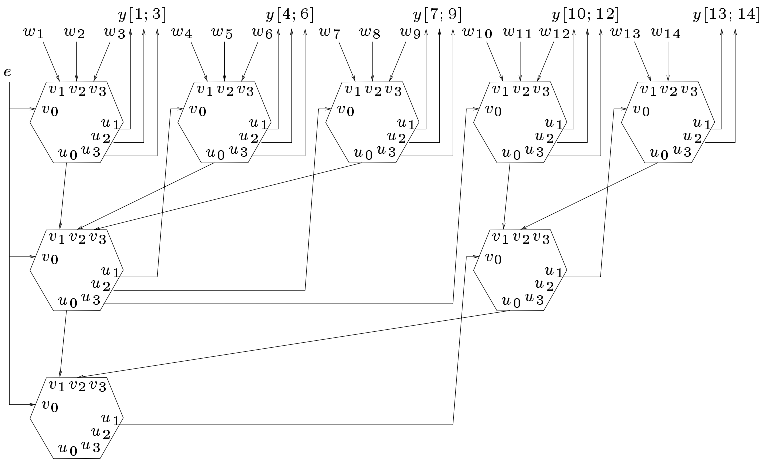

Figure 1: An expected delay optimal $\otimes_f^3$-circuit.

Using the standard unbounded fanin basis $\{\mathtt{OR}, \mathtt{AND}, \mathtt{NOT}\}$, an $\otimes_f^\ell$-gate can be replaced by a subcircuit of depth 3 and size $\mathrm{O}(\ell \cdot |\Sigma_I|^{\ell+1})$. Hence, the size of the resulting circuit is given by $\mathrm{O}(n \cdot |\Sigma_I|^{\ell+1})$. This implies:

Theorem 1 *For every $\ell \in \mathbb{N}$ and every $f \in \mathcal{F}_{\text{prefix}}^{\text{closed}}$, there exists a family of efficiently constructible circuits $C = C_1, C_2, \ldots$ over the standard basis $\{\mathtt{OR}, \mathtt{AND}, \mathtt{NOT}\}$ with fanin at most 2^ℓ that has size $s \le \mathrm{O}(n \cdot |\Sigma_I|^{\ell+1})$ and average delay*

$$\mathrm{etime}(C_n) \le \begin{cases} \mathrm{O}(1) & \text{if } f \in \mathcal{F}_{\text{str-conf}}^{\text{closed}} , \\ \mathrm{O}(1 + \ \mathrm{llog}\, n / \log \ell) & \text{if } f \in \mathcal{F}_{\text{conf}}^{\text{closed}} \cap \mathcal{F}_{\text{diff}}^{\text{closed}} , \\ \mathrm{O}(\log n / \log \ell) & \text{if } f \in \mathcal{F}_{\text{str-diff}}^{\text{closed}} . \end{cases}$$

Proof: To prove these bounds we will use two probability estimations shown in [Jak98, Jak99]:

Proposition 2 *For a prefix function $f \in \mathcal{F}_{\text{prefix}}$ and a uniformly distributed random string W of length $\ell \in \mathbb{N}$ there exists a constant $c \le \tau(f) + 2 \cdot \alpha(f) \cdot |\Sigma_I|^{\alpha(f)}$ such that for all $k \in \mathbb{N}$ with $c \cdot k \le \ell$ holds*

$$\Pr[\otimes(W[1; c \cdot k]) \notin \mathcal{Q}_f] \ \le\ 2^{-k} .$$

If further $f \in \mathcal{F}_{\text{conf}}$ *then it holds for all* $k \in \mathbb{N}$ *with* $c \cdot k \cdot \alpha(f) \leq \ell$

$$\Pr[\exists\, i \leq \ell - c \cdot k \cdot \alpha(f) \ :\ W[i; i + c \cdot k \cdot \alpha(f)] \textit{ is no confluence}] \quad \leq \quad \frac{\ell}{k} \cdot 2^{-k} \ .$$

Similarly to the case of bunded fanin [JRSW94] we can show that for any input $w \in \Sigma_I^n$ it holds

$$\text{time}_{C_n}(w) \leq 2 \cdot \max_{i+j \leq n} \ \{ \ \log j / \log \ell \mid w[i; i+j] \text{ is no confluence } \} + 2 \ .$$

Using the estimation given in Proposition 2 we get $\text{etime}(C_n) \leq \mathrm{O}(\text{llog}\, n / \log \ell)$ if f is confluent and $\text{etime}(C_n) \leq \mathrm{O}(\log n / \log \ell)$ otherwise. ∎

Choosing $\ell = \log^c n$ for $c > 0$ the size bound of Theorem 1 can be made arbitrarily close to linear depending on c. Thus we can state

Corollary 1 *For every* $\varepsilon > 0$ *and every function* $s = s(n)$ *in the range* $[O(n^{1+\varepsilon})..2^n]$, *there exists a family of efficiently constructible ub-circuits* $C = C_1, C_2, \ldots$ *for* $f \in \mathcal{F}_{\text{prefix}}^{\text{closed}}$ *having size* s *and average delay*

$$\text{etime}(C_n) \ \leq \ \begin{cases} \mathrm{O}(1) & \textit{if } f \in \mathcal{F}_{\text{conf}}^{\text{closed}} \ , \\ \mathrm{O}(\log n \ / \ \text{llog}\, s(n)) & \textit{if } f \in \mathcal{F}_{\text{str-diff}}^{\text{closed}} \ . \end{cases}$$

Summarizing the analysis, the difference in the average delay in these two cases is due to the different expected length of an interval of the input that does not contain a confluence. This length is logarithmic if f is confluent, otherwise of linear order.

5 Group Properties of Strictly Diffluent Functions

To investigate the average time complexity of prefix functions in the unbounded fanin circuit model in detail we have to discuss group-like properties of strict diffluences. Let f be a closed prefix function with domain Σ_I^* and $\langle \Sigma, \otimes \rangle$ be the corresponding groupoid. If $(\Sigma, \otimes)$ is a group then obviously f is strictly diffluent. In the following we will focus on some weaker properties that characterise strictly diffluent functions.

Definition 7 *For* $k \in \mathbb{N}$ *consider the* k*-dimensional space* Σ^k. *Its elements will be denoted by* $q^{[k]}$ *or* $\langle q_1, \ldots, q_k \rangle$, *where* $q_i \in \Sigma$. $q^{[k]}$ *will be called* **injektive** *if its components* q_i *are pairwise different. For* $w \in \Sigma_I^*$ *define a transition function on* Σ^k *by* $\otimes(\langle q_1, \ldots, q_k \rangle, w) \ := \ \langle \otimes_{q_1}(w), \ldots, \otimes_{q_k}(w) \rangle$. *This induces a* k*-dimensional graph* $\boldsymbol{G_f^k} := (\Sigma^k, E^k)$ *with* $E^k := \{(q^{[k]}, \otimes(q^{[k]}, x) \mid q^{[k]} \in \Sigma^k, \ x \in \Sigma_I \ \}$.

Let u, w *be strings in* Σ_I^*. u *will be called an* **inverse** *of* $(q^{[k]}, w)$ *if* $\otimes(\otimes(q^{[k]}, w), u) \ = \ q^{[k]}$. u *is a* **loop string** *of* $q^{[k]}$ *if it is an inverse of* $(q^{[k]}, \lambda)$, *that is* $\otimes(q^{[k]}, u) = q^{[k]}$. *For a* **shift string** u *of a vector* $\langle q_1, q_2, \ldots, q_k \rangle$ *we require that* $\otimes(\langle q_1, q_2, \ldots, q_k \rangle, u) \ = \ \langle q_2, \ldots, q_k, q_1 \rangle$.

We call the groupoid $\langle \Sigma, \otimes \rangle$ *for* f $\boldsymbol{\ell}$**-robust** *if for some* $k \geq 2$ *there exists a injective node* $q^{[k]} = \langle q_1, \ldots, q_k \rangle$ *in* G_f^k *such that*

1.) $q_1, \ldots, q_k \in \otimes_q(\Sigma_I^\ell)$ *for some* $q \in \Sigma$,

2.) $q^{[k]}$ *has a loop string and a shift string both of length* ℓ, *and*

3.) *for any* $w \in \Sigma_I^*$ *there exists a string* u *of length at most* ℓ *that is an inverse of* $(q^{[k]}, w)$, *and furthermore,* $|w\, u| \bmod \ell = 0$.

ℓ-robustness of a groupoid is closely related to the property of having a subgroup. Since there are ℓ-robust groupoids without a neutral element there are ℓ-robust groupoids which are not groups. But, it is easy to see that each group is a 1-robust groupoid.

To see that there are groupoids with a subgroup that are not ℓ-robust consider the confluent prefix function *parity with reset* $\otimes$ defined in the introduction. Note that the parity function without reset can trivially be embedded into this extension. Furthermore, $\langle \{0, 1, r\}, \otimes \rangle$ has a confluence r. But, $\langle \{0, 1, r\}, \otimes \rangle$ is not ℓ-robust for any $\ell \in \mathbb{N}$, i.e. there are no confluent ℓ-robust groupoids. On the other hand, each 1-robust groupoid contains a subgroup.

Theorem 2 *Every strictly diffluent closed prefix function is* ℓ*-robust for some* $\ell \in \mathbb{N}$.

Proof: In the following we will successively construct a sequence $\langle q_1, \ldots, q_k \rangle$ of states fulfilling the property ℓ-robustness for some $\ell \in \mathbb{N}$, that means $\langle q_1, \ldots, q_k \rangle \in V^k$ is an injective node of a closed component of G_f^k

with $k \geq 2$. Furthermore, $\langle q_1, \ldots, q_k\rangle$ is chosen such that there exists a shift and a loop string of the same length.

In [Jak98] it has been shown that for a function $f \in \mathcal{F}^{\text{closed}}_{\text{prefix}}$ each subfunction $\langle \Sigma, \otimes_q\rangle$ with $q \in \Sigma$ has the same period length $\pi(f) \leq |\Sigma|$. Choose $\boldsymbol{\ell} \geq |\Sigma|^2 + 2 \cdot |\Sigma|$ as the least multiple of $\pi(f)$ such that there exists a state q with $q \in R_{f,q}(\ell)$. Further, let $\boldsymbol{\mathcal{D}_q} \subset R_{f,q}(\ell)$ be the set of states which are pairwise strictly diffluent to q. Then it holds:

Lemma 1 *For every function $f \in \mathcal{F}^{\text{closed}}_{\text{str-diff}}$ and all $q_1 \in \Sigma$ with $q_1 \in R_{f,q_1}(\ell)$ there exists $q_2 \in \mathcal{D}_{q_1}$ and two strings $w_{\text{loop}}, w_{\text{shift}} \in \Sigma_I^{j\cdot\ell}$ for some $j \in \mathbb{N}$ such that $q_1 = \otimes_{q_1}(w_{\text{loop}})$, $q_2 = \otimes_{q_2}(w_{\text{loop}})$ and $q_2 = \otimes_{q_1}(w_{\text{shift}})$. Further, q_2 can be chosen such that for any $w \in \Sigma_I^*$ there exists a inverse string of $(\langle q_1, q_2\rangle, w)$.*

Proof: For $w_1 \in \Sigma_I^\ell$ with $q_1 = \otimes_{q_1}(w_1)$ let $\mathcal{D}_{w_1} \subseteq \mathcal{D}_{q_1}$ be the set of states q_2 such that w_1^c maps q_2 to itself for some $c \geq 1$. From the definition of strict diffluence and the set $\mathcal{D}_{q_1}$ it follows that $\mathcal{D}_{w_1} \neq \emptyset$. Furthermore, for $q \in \mathcal{D}_{w_1}$ we choose the value c_q minimal such that $q = \otimes_q(w_1^{c_q})$. Let $\boldsymbol{\xi}$ be the least common multiple of all these values c_q and $|\Sigma|$.

For a string $w \in \Sigma_I^*$ with $q_1 = \otimes_{q_1}(w)$ choose $u \in \Sigma_I^*$ such that $q_1 = \otimes_{q_1}(wu)$ and $|wu| \bmod \ell = 0$. Since each element $q_2 \in \mathcal{D}_{w_1}$ is strictly diffluent to q_1 it holds that $\otimes_{q_2}(wuw_1^{|\Sigma|}) \in \mathcal{D}_{w_1}$. So we can construct a directed graph $G_{w_1} = (\mathcal{D}_{w_1}, E_{w_1})$ where

$$(q, q') \in E_{w_1} \qquad \Longleftrightarrow \qquad \exists i \in \mathbb{N} \quad \exists w \in \Sigma_I^{i\cdot\ell}: \quad q' = \otimes_q(w) \ \text{ and } \ q_1 = \otimes_{q_1}(w) \ .$$

Let q be a node of a closed component of G_{w_1} and $\boldsymbol{\mathcal{D}_{w_1,q}}$ be the set of states of these component. Then for any string w with $|w| \bmod \ell = 0$ and $q_1 = \otimes_{q_1}(w)$ it holds $\otimes_q(ww_1^{|\Sigma|}) \in \mathcal{D}_{w_1,q}$.

Note that a node $\langle q_1, q_2\rangle$ of G_f^2 is an element of a closed component of G_f^2 if for all $w \in \Sigma_I^*$ there exists an inverse string for $\langle q_1, q_2\rangle$ and w. Hence, one can find an inverse u for q and $ww_1^{|\Sigma|}$ with $|u| \bmod \ell = 0$ and $q_1 = \otimes_{q_1}(u)$. The second part of the claim follows directly.

Let $q' \in \mathcal{D}_{w_1,q} \subseteq \mathcal{D}_{w_1} \subseteq \mathcal{D}_{q_1} \subset R_{f,q_1}(\ell)$ and $w_2 \in \Sigma_I^\ell$ with $q' = \otimes_{q_1}(w_2)$. Furthermore, define $w_{\text{loop}} := w_1^\xi$, $w_{\text{shift}} := w_2 w_1^{\xi-1}$ and $q_2 := \otimes_{q_1}(w_{\text{shift}})$. Hence, the first part of the claim follows directly from the definition of $\mathcal{D}_{w_1,q_1}$. ∎

Using q_1, q_2 and the strings $w_{\text{loop}} = w_1^\xi$ and w_{shift} constructed above we can successively generate a sequence $\langle q_1, q_2, q_3, \ldots, q_m\rangle$ of elements of Σ as well as two sequences of strings $w^{[1]}_{\text{loop}}, \ldots, w^{[m]}_{\text{loop}}$ and $w^{[1]}_{\text{shift}}, \ldots, w^{[m]}_{\text{shift}}$ such that for all $j \in [1;m]$ the following properties holds:

1) For all $i \in [1;j]$ $\quad q_i = \otimes_{q_i}(w^{[j]}_{\text{loop}})$ and for all $i \in [1;j-1]$ $\quad q_{i+1} = \otimes_{q_i}(w^{[j]}_{\text{shift}})$.
2) For each $\langle q_1, q_2, q_3, \ldots, q_m\rangle \in V_f^m$ and each $w \in \Sigma_I^*$ there exists an inverse u with $|wu| \bmod \ell = 0$.
3) There exists $r \in [1;m]$ with $q_r = \otimes_{q_m}(w^{[m]}_{\text{shift}})$.

This can be shown analogously to the construction given in Lemma 1. ∎

Definition 8 *For a function $f \in \mathcal{F}^{\text{closed}}_{\text{str-diff}}$ define $\boldsymbol{\ell(f)}$ as the minimal $\ell \geq |\Sigma|^2 + 2 \cdot |\Sigma|$ such that f is ℓ-robust, i.e. there exists a sequence $\langle q_1, \ldots, q_m\rangle$ as well as a loop and a shift string $w^{[m]}_{\text{loop}}, w^{[m]}_{\text{shift}} \in \Sigma_I^{\ell(f)}$ as constructed in the previous Theorem.*

For a prefix function $f \in \mathcal{F}^{\text{closed}}_{\text{str-diff}}$ and an input $w \in \Sigma_I^*$ define $\boldsymbol{f_{\ell(f)}(w)} := f(w)[1 \cdot \ell(f)]f(w)[2 \cdot \ell(f)] \ldots$ as the subfunction of f computing only the values at output positions that are multiples of $\ell(f)$.

Let $\boldsymbol{\mathcal{F}_{\circ-\text{diff}}}$ be the set of all 1-robust prefix functions $f \in \mathcal{F}^{\text{closed}}_{\text{str-diff}}$ with a complete transition graph (i.e. for all pairs $q, q' \in \Sigma$ there exists a symbol $x \in \Sigma_I$ such that $q = \otimes_{q'}(x)$).

The following properties can easily be seen:

1.) If the underlying groupoid of f is a semigroup it holds $\ell(f) = |\Sigma|^2 + 2 \cdot |\Sigma|$.
2.) For the restriction of prefix functions f that are defined by an associative operator $\otimes$, $f_{\ell(f)}$ belongs to $\mathcal{F}_{\circ-\text{diff}}$.
3.) $f_{\ell(f)}$ is equivalent to a prefix function over the input alphabet $\Sigma_I' := \Sigma_I^{\ell(f)}$. Furthermore, $f_{\ell(f)}$ is 1-robust.

A lower bound on the expected delay of a ub-circuits computing $f_{\ell(f)}$ implies the same lower bound for the expected delay of f. Therefore, we can restrict our analysis to 1-robust functions.

6 Error Bounds for Circuits Computing Strictly Diffluent Prefix Functions over Arbitrary Alphabets

To obtain lower bounds on the average delay, we will estimate the maximal number of inputs for which an ub-circuit can give correct results for a given prefix function $f \in \mathcal{F}_{\circ-\text{diff}}$. The heart of the probabilistic restriction method for ub-circuits is the efficient interchange of conjunctions and disjunctions, thus it can only be applied in a Boolean setting. As already discussed, nonbinary groupoids Σ have to be coded by binary strings and appropriate subcircuits. If $|\Sigma|$ is not a power of 2, let's say $2^{\kappa-1} < |\Sigma| < 2^{\kappa}$ for some natural number κ, then certain binary strings of length κ will not be used as codewords.

However, in the course of fixing certain inputs bits by probabilistic selection, it may happen that strings are generated that are no codewords. One cannot simply use them as alternative codewords for some elements of Σ because this would bias the distribution.

Generalizing the calculations of Håstad [Hast86b] for depth d Boolean ub-circuit of polynomial size with $d \in$ o(log / llog) to alphabets of size 2^{κ}, the number of inputs of length n, for which such a circuit does not compute the correct result, gets bounded by

$$2^{\kappa\, n} \cdot \left(\frac{1}{2^{2\kappa}} - 2^{-\Omega((n/\log n)^{1/d})} \right) \;\geq\; \frac{1}{2^{2\kappa+1}}\, 2^{\kappa\, n} \qquad \text{for large } n .$$

However, only $|\Sigma|^n$ strings are relevant for the prefix function. Thus, due to the loss in the coding rate if $|\Sigma| < 2^{\kappa}$, this number is much smaller than the remaing set of inputs, for which the circuit works correctly, which is of size $\Omega(2^{\kappa\, n})$.

Hence, this simple generalization does not provide good lower bounds for general alphabets. Therefore, a more sophisticated approach is needed that simulates a probabilistic restriction for an alphabet Σ in a Boolean setting.

Definition 9 *For $0 \leq p \leq 1$, the* **probabilistic restriction** *over the alphabet* $\boldsymbol{\Sigma_I}$, *denoted by* $\boldsymbol{\rho_{[p,\Sigma_I]}}$, *independently sets the input variables B_i as follows:*

$$\Pr[\rho_{[p,\Sigma_I]}(B_i) = \star] = p \qquad \textit{and} \qquad \forall x \in \Sigma_I \;:\; \Pr[\rho_{[p,\Sigma_I]}(B_i) = x] = \frac{1-p}{|\Sigma_I|} .$$

In the following we will restrict ourself to functions $f' : \Sigma_I^n \to \{0,1\}$ that for a given prefix function f, an input $w \in \Sigma_I^n$, and an symbol $x \in \Sigma$ decide whether $f(w)[n] = y$, that means whether the last value of $f(w)$ is y or not.

For a circuit C_n with n inputs and a string $w \in (\Sigma_I \cup \{\star\})^n$ define $\boldsymbol{C\lceil_w}$ as the simplified circuit (resp. the function computed by the circuit) for which all inputs v_i with $w[i] \neq \star$ are determined by $w[i]$. To consider arbitrary input alphabets Σ_I we will use special input gates, denoted by $\boldsymbol{[B \in Z]}$ for an input variable B and a subset $Z \subseteq \Sigma_I$ that evaluates to 1 iff $B \in Z$. A conjunction, resp. disjunction of the variables $B_1, \ldots, B_n$ is a term of type

$$\boldsymbol{K_i} \;:=\; \bigwedge_{j \in \{1,\ldots,|K_i|\}} [B_{\pi_i(j)} \in Z_{i,j}], \qquad \text{resp.} \quad \boldsymbol{D_i} \;:=\; \bigvee_{j \in \{1,\ldots,|D_i|\}} [B_{\pi_i(j)} \in Z_{i,j}]$$

with $\boldsymbol{Z_{i,j}} \subset \Sigma_I$. A conjunction K_i is called a **conjunctive min-term** if

$$\forall w \in \Sigma_I^n \;:\; K_i(w) = 1 \;\to\; f(w) = 1 \quad \text{and} \quad \forall\, k \in [1; |K_i|] \;\; \exists\, w \in \Sigma_I^n \;:\; f(w) = 0 \;\wedge\; K_{i,k}(w) = 1 ,$$

where $K_{i,k} \;:=\; \bigwedge_{j \in \{1,\ldots,|K_i|\} \setminus \{k\}} [B_{\pi_i(j)} \in Z_{i,j}]$. A **disjunctive min-term** is defined analogously. Further, $\boldsymbol{\min(C)}$ denotes the maximum number of variables of a conjunctive min-term of C. Generalizing the analysis in [Hast86b] it can be shown:

Lemma 2 *Let $C = \bigwedge_{i=1}^{k} D_i$ where the D_i are disjunctions of at most t variables. Further, let $F : \Sigma_I^* \to \{0,1\}$ be an arbitrary function and $\rho = \rho_{[p,\Sigma_I]}$ a probabilistic restriction. Then it holds*

$$\Pr[\min(C\lceil_\rho) \geq s \mid F\lceil_\rho \equiv 1] \;\leq\; \alpha_\rho^s ,$$

where α_ρ is the unique positive solution of

$$\left(1 + \frac{|\Sigma_I|^2 \cdot p}{1 + (|\Sigma_I| - 1) \cdot p} \cdot \frac{1}{\alpha_\rho}\right)^t \;=\; \left(1 + \frac{|\Sigma_I| \cdot p}{1 + (|\Sigma_I| - 1) \cdot p} \cdot \frac{1}{\alpha_\rho}\right)^t + 1 .$$

Investigating $F\lceil_\rho \equiv 0$ instead of $F\lceil_\rho \equiv 1$ the same technique can be used to transform a disjunction of conjunctions into a conjunction of disjunctions.

Lemma 3 *The value of α_ρ can be bounded by the following inequality:*

$$\frac{t \cdot (|\Sigma_I| \cdot \ln 2 - 1) \cdot |\Sigma_I| \cdot p}{(1 + (|\Sigma_I| - 1) \cdot p) \cdot \ln 2} \;<\; \alpha_\rho \;\leq\; \frac{t \cdot |\Sigma_I|^2 \cdot p}{(1 + (|\Sigma_I| - 1) \cdot p) \cdot \ln 2} \;.$$

For a gate v of C define **accept(v)** as the set of inputs $w \in \Sigma_I^n$ such that $\mathrm{res}_v(w) = 1$, and **accept(C)** $:=$ accept(v) where v denotes the output gate of C. If we allow arbitrary conjunctions for $C\lceil_\rho$ (not only conjunctive min-terms) we can improve Lemma 2 to:

Lemma 4 *Let $C = \bigwedge_{i=1}^k D_i$ and ρ as above. Then the probability that $C\lceil_\rho$ can be represented by a disjunction of conjunctions K_j with maximal degree s is at least $1 - \alpha_\rho^s$. Furthermore, this conjunction can be chosen such that the sets* accept(K_i) *are pairwise disjoint.*

Exploiting the robustness of strictly diffluent prefix functions f one can show an upper bound on the maximal correctness of a d-depth s-size bounded ub-circuit C for inputs of length n. Let us denote this value by $\boldsymbol{\mathcal{E}(f, d, s, n)}$, i.e. there exists at most $\mathcal{E}(f, d, s, n) \cdot |\Sigma_I|^n$ inputs $w \in \Sigma_I^n$ with $C(w) = f(w)$.
For a function $f \in \mathcal{F}_{\circ-\mathrm{diff}}$ define $\boldsymbol{f_{\mathbf{last}}} : \Sigma_I^* \to \Sigma$ as the rightmost output symbol of f, that means $f_{\mathbf{last}}(w) := f(w)[|w|]$. Let $\mathcal{F}_{\circ-\mathbf{diff}}^{\mathbf{last}}$ denote the set of all such functions f_{last}. Since $\mathcal{E}(f, d, s, n) \leq \mathcal{E}(f_{\mathrm{last}}, d, s, n)$ for $f \in \mathcal{F}_{\circ-\mathrm{diff}}$ we can resctrict our investigations to the set $\mathcal{F}_{\circ-\mathrm{diff}}^{\mathrm{last}}$.

Theorem 3 *For $f \in \mathcal{F}_{\circ-\mathrm{diff}}^{\mathrm{last}}$ it holds*

$$\mathcal{E}(f, d, s, n) \;\leq\; \frac{|\Sigma_I|^2 - 1}{|\Sigma_I|^2} + \Delta \qquad \textit{where} \qquad \Delta \;:=\; \begin{cases} |\Sigma_I|^{-\Omega(n/\log_{|\Sigma_I|} s)} & \textit{if } d = 2, \\ |\Sigma_I|^{-\Omega(n/(\log_{|\Sigma_I|} s)^{d-1})} & \textit{if } \log_{|\Sigma_I|} s \geq n^{1/d}, \\ |\Sigma_I|^{-\Omega((n/\log_{|\Sigma_I|} s)^{1/d})} & \textit{if } \log_{|\Sigma_I|} s < n^{1/d}. \end{cases}$$

7 Lower Bounds on the Average Delay

For a function $f \in \mathcal{F}_{\mathrm{prefix}}^{\mathrm{closed}}$ the value $\mathcal{E}(f, d, s, n)$ allows us to derive a lower bound on the expected delay of an ub-circuit for f.

Theorem 4 *For $f \in \mathcal{F}_{\text{str-diff}}^{\mathrm{closed}}$ it holds for any size bound s*

$$\mathrm{etime}(\mathrm{UbCir}(f_n, s)) \;\geq\; \Omega\left(\frac{\log n}{\operatorname{llog} s}\right) \;.$$

Proof: We will use a technique introduced in [JRS99]. Any family of circuits C_n of size $s = s(n)$ computing f within an expected delay of $t = t(n)$ can be transformed into a depth t-bounded family of circuits C_n' of size at most quadratic in s computing f_n with an error rate tending to 0. This implies $1 - \mathcal{E}(f, t, s, n) \in o(1)$. Considering the value of Δ this means that we have to choose a bound t for the expected delay such that Δ tends to $1/|\Sigma_I|^2$ for growing n. This requires for all n large enough $\Delta > 1/|\Sigma_I|^3$.
According to the different cases in Theorem 3 we can conclude:
(1) If $t = 2$ then $3 \geq \Omega(n/\log_{|\Sigma_I|} s)$, that means $s \geq \exp \Omega(n)$.
(2) Now, it is required that $3 \geq \Omega(n/(\log_{|\Sigma_I|} s)^{t-1})$, or equivalently $(\log_{|\Sigma_I|} s)^{t-1} \geq \Omega(n)$. This imlies

$$t \geq \frac{\log n}{\operatorname{llog} s - \operatorname{llog} |\Sigma_I|} - O(1) \;.$$

(3) The last case cannot occur since it does not satisfy the correctness requirements above. The inequality $3 \geq \Omega((n/\log_{|\Sigma_I|} s)^{1/t})$ together with the condition $\log_{|\Sigma_I|} s < n^{1/t}$ imply

$$\Omega(n^{1/t}) \;\leq\; (\log_{|\Sigma_I|} s)^{1/t} \;\leq\; n^{1/t^2} \;, \qquad \text{that means} \qquad n^{1/t^2 - 1/t} \geq \Omega(1) \;.$$

Since $t > 1$ this yields a contradiction. ∎
The Main Theorem follows by combining the last theorem with the upper bounds shown in section 4.

8 Conclusions

This paper gives optimal lower and upper bounds for the average-case complexity of arbitrary prefix functions in the unbounded fanin circuit model. For the expected delay we have shown the same trade-off with respect to the circuits size for functions like `PARITY` as it has been known for the worst-case complexity. However, there are other classes of functions like *parity with reset* with a huge gap between worst and averge case complexity. Finally, our results show that prefix functions split into only two different complexity classes depending on the algebraic structure of the underlying groupoid: the easy ones with almost linear size and constant delay, and the hard ones with the same complexity as `PARITY`.

References

[Bab87] L. Babai, *Random Oracle Separates PSPACE from the Polynomial-Time Hierarchy,* IPL 26, 1987, 51-53.

[BiPr89] G. Bilardi, F. Preparata, *Size-Time Complexity of Boolean Networks for Prefix Computations,* J. of the ACM 36, 1989, 362-382.

[BiPr90] G. Bilardi, F. Preparata, *Characterization of Associative Operations with Prefix Circuits of Constant Depth and Linear Size,* SIAM J. Comput 19, 1990, 246-255.

[Cai89] J. Cai, *With Probability One, a Random Oracle Separates PSPACE from the Polynomial-Time Hierarchy,* J. CSS 38, 1989, 68-85.

[ChFL83a] A. Chandra, S. Fortune, R. Lipton, *Unbounded Fan-in Circuits and Associative Functions*, Proc. 15. SToC'83, 1983, 52-60.

[ChFL83b] A. Chandra, S. Fortune, R. Lipton, *Lower Bounds for Constant Depth Circuits for Prefix Problems*, Proc. 10. ICALP'83, 1983, 109-116.

[FuSS84] M. Furst, J. B. Saxe, M. Sipser, *Parity, Circuits, and the Polynomial-Time Hierarchy*, Proc. 22. FoCS'81, 1981, 260-270, and Math. Systems Theory 17, 1984, 13-27.

[Hast86a] J. Håstad, *Almost Optimal Lower Bounds for Small Depth Circuits*, Proc. 18. SToC'86, 1986, 6-20.

[Hast86b] J. Håstad, *Computational Limitations for Small-Depth Circuits*, ACM Doctoral Dissertation Award 1986, MIT Press, 1986.

[Jak98] A. Jakoby, *Die Komplexität von Präfixfunktionen bezüglich ihres mittleren Zeitverhaltens*, Dissertation, Med. Universität zu Lübeck, 1998.

[Jak99] A. Jakoby, *The Average Time Complexity to Compute Prefix Functions in Processor Networks*, Proc. 16. STACS'99, 1999, 78-89.

[JRS94] A. Jakoby, R. Reischuk, C. Schindelhauer, *Circuit Complexity: from the Worst Case to the Average Case*, Proc. 26. ACM Symposium on the Theory of Computing, 1994, 58-67.

[JRS99] A. Jakoby, R. Reischuk, C. Schindelhauer, *Malign Distributions for Average Case Circuit Complexity*, Information & Computation 150, 1999, 187-208.

[JRSW94] A. Jakoby, R. Reischuk, C. Schindelhauer, S. Weis, *The Average Case Complexity of the Parallel Prefix Problem*, Proc. 21. ICALP'94, 1994, 593-604.

[Kra78] V. Krapchenko, *Depth and Delay in a Network,* Soviet Math. Dokl. 19, 1978, 1006-1009.

[LaFi80] R. Ladner, M. Fischer, *Parallel Prefix Computation*, J. ACM 27, 1980, 831-838.

[Smo87] R. Smolensky, *Algebraic Methods in the Theory of Lower Bounds for Boolean Circuit Complexity,* Proc. 19. SToC, 1987, 77-82.

[Yao85] A. Yao, *Separating the Polynomial Time Hierarchy by Oracles,* Proc. 26. FoCS, 1985, 1-10.

9 Appendix

Proof of Theorem 1: For easier notion we assume that $\langle\Sigma,\otimes\rangle$ is a semigroup, i.e. that $\otimes$ is associative with a left neutral element e, i.e. $\otimes(ew)=\otimes(w)$ for all $w\in\Sigma_I^*$.
In the following we will construct an $O(n/\ell)$ size bounded circuit C_n for inputs of length n computing a prefix function f using only $\otimes_f^\ell$-gates where ℓ may depend on the input length. For $\ell\in\mathbb{N}$ define a $\otimes_f^\ell$-gate with inputs $v_0,\ldots,v_\ell$ and outputs $u_0,\ldots,u_\ell$ by

$$\mathrm{res}_{u_0}(w) \;:=\; (\cdots((\mathrm{res}_{v_1}(w)\otimes\mathrm{res}_{v_2}(w))\otimes\mathrm{res}_{v_3}(w))\otimes\cdots)\otimes\mathrm{res}_{v_\ell}(w)$$

and for all $i\in\{1,\ldots,\ell\}$ $\quad \mathrm{res}_{u_i}(w) \;:=\; (\cdots((\mathrm{res}_{v_0}(w)\otimes\mathrm{res}_{v_1}(w))\otimes\mathrm{res}_{v_2}(w))\otimes\cdots)\otimes\mathrm{res}_{v_i}(w)$

u_0 combines all inputs except v_0. The outputs u_i with $i\geq 1$ gives the parallel prefix function over $\langle\Sigma_f,\otimes_f\rangle$ for the input $v_1,\ldots,v_\ell$ where the *starting state* of f is given by v_0.
The basic structure of C_n is given by an almost complete ℓ-ary in-tree $G_T:=(V_T,E_T)$ with

$$V_T \;:=\; \{\; g_{i,j}\mid 1\leq i\leq\lceil\log n/\log\ell\rceil \text{ and } 1\leq j\leq\lceil n/\ell^i\rceil\;\}$$

where all paths from a leaf to the root have length $\log n/\log\ell$. Furthermore, every internal node except the ones on the rightmost path have indegree exactly ℓ, while those on the rightmost path have indegree at most ℓ. That means,

$$E_T \;:=\; \{\; (g_{i-1,(j-1)\cdot\ell+k},g_{i,j})\mid 1\leq k\leq\ell \;\text{ and }\; g_{i-1,(j-1)\cdot\ell+k},g_{i,j}\in V_T\;\}\;.$$

To connect two gates $g_{i-1,(j-1)\cdot\ell+k},g_{i,j}$ with $(g_{i-1,(j-1)\cdot\ell+k},g_{i,j})\in E_T$ we will draw a wire from the output u_0 of $g_{i-1,(j-1)\cdot\ell+k}$ to the input v_k of $g_{i,j}$. Furthermore, we use the inputs $v_1,\ldots,v_\ell$ of the leafs $g_{1,i}$ as the inputs of the circuit.
Our construction will be completed by adding edges $(g_{i,j},g_{i-1,(j-1)\cdot\ell+k+1})$ with $1\leq k\leq\ell$, $2\leq i\leq\log n/\log\ell$, and $1\leq j\leq n/\ell^i$ to G_T. To connect $g_{i,j}$ with $g_{i-1,(j-1)\cdot\ell+k+1}$ we will use a wire from the output u_k of $g_{i,j}$ to the input v_0 of $g_{i-1,(j-1)\cdot\ell+k+1}$.
Similarly to the case of bunded fanin [JRSW94] we can show that for any input $w\in\Sigma_I^n$ it holds

$$\mathrm{time}_{C_n}(w)\leq 2\cdot\max_{i+j\leq n}\;\{\;\log j/\log\ell\mid w[i;i+j] \text{ is no confluence }\}+2\;.$$

Using the estimation given in Proposition 2 we get

$$\mathrm{etime}(C_n)\leq\begin{cases}\mathrm{O}(\mathrm{llog}\,n/\log\ell) & \text{if } f\in\mathcal{F}_{\text{conf}}^{\text{closed}}\\ \mathrm{O}(\log n/\log\ell) & \text{if } f\in\mathcal{F}_{\text{str-diff}}^{\text{closed}}\;.\end{cases}$$

Furthermore, if we use the standard basis $\{\mathtt{OR},\mathtt{AND},\mathtt{NOT}\}$ of fanin $|\Sigma_I|^\ell$, any $\otimes_f^\ell$-gate can be replaced by a subcircuit of depth 3 and size $\Theta(\ell\cdot|\Sigma_I|^{\ell+1})$. Hence, the size of the resulting circuit is given by $\Theta(n\cdot|\Sigma_I|^{\ell+1})$, whereas the upper bound for the expected time of the resulting circuit is given by the estimation above. ∎

Proof of Theorem 2: To prove that f is ℓ-robust for some multiple of $\pi(f)$ we have to show that there exists a constant $k\leq|\Sigma|$, a sequence of states $q_0,\ldots,q_{k-1}\in\Sigma$, and two strings $w_{\text{shift}},w_{\text{loop}}\in\Sigma_I^\ell$ such that:

1. for each q_i holds $q_{(i+1)\bmod k}=\otimes_{q_i}(w_{\text{shift}})$ and $q_i=\otimes_{q_i}(w_{\text{loop}})$ and
2. for all strings $u\in\Sigma_I^*$ there exists an inverse string $w\in\Sigma_I^*$ with $|uw|\bmod\ell=0$, that means that for all q_i holds $q_i=\otimes_{q_i}(uw)$.

Using q_1,q_2 as well as the strings $w_{\text{loop}}=w_1^\xi$ and w_{shift} constructed in Lemma 1 we can successively generate a sequence $\langle q_1,q_2,q_3,\ldots,q_m\rangle$ of elements of Σ as well as two sequences of strings $w_{\text{loop}}^{[1]},\ldots,w_{\text{loop}}^{[m]}$ and $w_{\text{shift}}^{[1]},\ldots,w_{\text{shift}}^{[m]}$ such that for all $j\in[1;m]$ the following properties holds:
1) For all $i\in[1;j]$ $\quad q_i=\otimes_{q_i}(w_{\text{loop}}^{[j]})$.
2) For all $i\in[1;j-1]$ $\quad q_{i+1}=\otimes_{q_i}(w_{\text{shift}}^{[j]})$.
3) For each $\langle q_1,q_2,q_3,\ldots,q_m\rangle\in V_f^m$ and each $w\in\Sigma_I^*$ there exists an inverse u with $|wu|\bmod\ell=0$.
4) There exists $r\in[1;m]$ with $q_r=\otimes_{q_m}(w_{\text{shift}}^{[m]})$.

Figure 2: Sequence $\langle q_1, q_2, q_3, q_4, q_5, q_6\rangle$ of a closed component of G_f^k.

This can be shown analogously to the construction given in the proof of Lemma 1. Figure 2 illustrates the first three conditions of the sequence $\langle q_1, q_2, q_3, q_4, q_5, q_6\rangle$ with $r = 3$.
Assume that $q_1, \ldots, q_i$, $w_{\text{loop}}^{[i-1]}$, and $w_{\text{shift}}^{[i-1]}$ with $i \le m$ fulfil the conditions 1 to 3 of the sequence $\text{Oppo}(q_1, w_{\text{loop}}, w_{\text{shift}})$. Hence, the states $q_1, \ldots, q_i$ are pairwise strictly diffluent. The succeeding element q_{i+1} as well as two strings $w_{\text{loop}}^{[i]}$ and $w_{\text{shift}}^{[i]}$ are constructed as follows:
Let $\mathcal{D}_{w_1}$, ξ, and ℓ be chosen as in Lemma 1 and define $v_{i+1} := \otimes_{q_i}(w_{\text{shift}}^{[i-1]})$. Note that q_{i-1} and q_i are strictly diffluent and therefore q_i and v_{i+1} too. Investigating the pair q_i, v_{i+1} we have to distinguish the following two cases

1) There exists an element q_r of the input sequence with $r < i$ such that q_r and v_{i+1} are confluent.

2) All states q_j with $j \le i$ are strictly diffluent to v_{i+1}.

In the first case let w be a confluence of q_r and v_{i+1}. Further, let $u \in \Sigma_I^*$ be an inverse string of $\langle q_1, \ldots, q_i\rangle$ and w with $|wu| \bmod \ell = 0$. Then the sequence $\langle q_1, \ldots, q_i\rangle$, as well as

$$w_{\text{loop}}^{[i]} \;:=\; w_{\text{loop}}^{[i-1]} wu, \qquad \text{and} \qquad w_{\text{shift}}^{[i]} \;:=\; w_{\text{shift}}^{[i-1]} wu$$

fulfils the conditions of $\text{Oppo}(q_1, w_{\text{loop}}, w_{\text{shift}})$ where $m := i$.
In the second case we choose q_{i+1}, $w_{\text{loop}}^{[i]}$, and $w_{\text{shift}}^{[i]}$ analogously to Lemma 1:
For the sequence $\langle q_1, \ldots, q_i\rangle$ let $\boldsymbol{Q_i}$ be the set of elements $q \in R_{f,q_1}(\ell) \setminus \{q_1, \ldots, q_i\}$ with $q = \otimes_q(w_1^\xi)$ which are pairwise strictly diffluent to the elements of $\{q_1, \ldots, q_i\}$. Since the pair q_j, v_{i+1} is strictly diffluent for all $j \in [1; i]$ it holds

$$v'_{i+1} \;:=\; \otimes_{v_{i+1}}(w_1^\xi) \;\in\; Q_i \;.$$

Let $u_{\text{loop}}^{[i]} := w_{\text{loop}}^{[i-1]} w_1^\xi$ and $u_{\text{shift}}^{[i]} := w_{\text{shift}}^{[i-1]} w_1^\xi$. To find a successor of v'_{i+1} that fulfills the third condition of $\text{Oppo}(q_1, w_{\text{loop}}, w_{\text{shift}})$ we define a graph $G_i := (Q_i, E_i)$ with

$$(q, q') \in E_i \qquad \Longleftrightarrow \qquad \exists j \in \mathbb{N}\ \exists w \in \Sigma_I^{j\cdot\ell} :\ q' = \otimes_q(w) \ \text{ and }\ \forall k \le i :\ q_k = \otimes_{q_k}(w) \;.$$

Let q be an element of a closed component $\boldsymbol{\mathcal{Q}_i}$ of G_i. Furthermore, let $\boldsymbol{w_{\text{in}}}$ be a string of length ℓ with

$$q = \otimes_{v'_{i+1}}(w_{\text{in}}) \qquad \text{and} \qquad \forall j \in [1; i] :\ q_j = \otimes_{q_j}(w_{\text{in}}) \;.$$

Note that the construction of G_i implies the existence of such a state q and a string w_{in}. Choosing

$$w_{\text{loop}}^{[i]} \;:=\; u_{\text{loop}}^{[i]} w_1^\xi, \qquad w_{\text{shift}}^{[i]} \;:=\; u_{\text{shift}}^{[i]} w_{\text{in}} w_1^{\xi-1}, \qquad \text{and} \qquad q_{i+1} := \otimes_q(w_1^{\xi-1}) = \otimes_{q_i}(w_{\text{shift}}^{[i]})$$

it holds $|w_{\text{loop}}^{[i]}| = |w_{\text{shift}}^{[i]}|$,

$$\forall j \in [1; i+1] \ :\ q_j = \otimes_{q_j}(w_{\text{loop}}^{[i]}), \qquad \text{and} \qquad \forall j \in [1; i] \ :\ q_{j+1} = \otimes_{q_j}(w_{\text{shift}}^{[i]}) \;.$$

Finally we have to show that for each string w there exists a inverse string u for $\langle q_1, \ldots, q_{i+1}\rangle \in V_f^{i+1}$ and w with $|wu| \bmod \ell = 0$.

From the third condition of the sequence $q_1, \dots, q_i$ it follows that for each string w there exists an inverse string u for $\langle q_1, \dots, q_i\rangle$ and w with $|wu| \bmod \ell = 0$. Let $q := \otimes_{q_{i+1}}(wuw_1^\xi)$. Then it follows from the definition of G_i that q is an element of $\mathcal{Q}_i$. Hence, there exists a string w' with $q_{i+1} = \otimes_q(w')$ where

$$|w'| \bmod \ell = 0 \qquad \text{and} \qquad \forall j \in [1;i]: \ q_j = \otimes_{q_j}(w')\ ,$$

i.e. for all $j \in [1; i+1]$ it holds $q_j = \otimes_{q_j}(wuw_1^\xi w')$.
These properties imply Theorem 2. ∎

Proof of Lemma 2: Let $\boldsymbol{\mathcal{MT}(f)}$ be the set of all conjunctive min-terms of f and define $\boldsymbol{\mathcal{MT}(C)}$ as the set of all conjunctive min-terms of the corresponding function. For easier notion let $\boldsymbol{\rho} := \rho_{[p,\Sigma_I]}$.
The proof follows by a induction over the number of disjunctions k. Note that the claim holds for $k \leq 1$. If $k \geq 2$ we get

$$\begin{aligned}\Pr[\min(C\lceil_\rho) \geq s \mid F\lceil_\rho \equiv 1] \leq \max\{\ & \Pr[\min(C\lceil_\rho) \geq s \mid F\lceil_\rho \equiv 1 \ \wedge\ D_1\lceil_\rho \equiv 1], \\ & \Pr[\min(C\lceil_\rho) \geq s \mid F\lceil_\rho \equiv 1 \ \wedge\ D_1\lceil_\rho \not\equiv 1]\ \}\ .\end{aligned} \tag{1}$$

In the first case the claim follows directly from the induction.
If $D_1\lceil_\rho \equiv 0$ it holds that $C\lceil_\rho \equiv 0$ and therefore $\Pr[\min(C\lceil_\rho) \geq s \mid D_1\lceil_\rho \equiv 0] = 0$. So we have to investigate the case that $D_1\lceil_\rho \not\equiv 0$ as well as $D_1\lceil_\rho \not\equiv 1$. Hence, the restriction ρ does not determine the value of D_1.
If there exists a satisfying assignment for a conjunctive min-term $K_j := \bigwedge_i [B_{\pi_j(i)} \in Z_{j,i}] \in \mathcal{MT}(C\lceil_\rho)$ then there exists at least one test $[B_{\pi_j(i)} \in Z_{j,i}]$ in K_j and one test $[B_{\pi_j(i)} \in Z'_{j,i}]$ in D_1 such that $Z_{j,i} \subseteq Z'_{j,i}$.
Let T denote the set of variables used in disjunction $D_1 = \bigvee_{i\in T}[B_i \in Z_i]$. Note that for any variable B_i there is at most one test $[B_i \in Z_i]$ pro clause. W.l.o.g. assume that $T = [1; |T|]$. Hence, there exists a partition $\boldsymbol{\rho_1\rho_2}$ of ρ such that ρ_1 gives the restrictions of the variables of D_1 and ρ_2 the restrictions of the variables B_i with $i \notin T$. It follows, that the second term of estimation 1 is equal to

$$\Pr[\min(C\lceil_\rho) \geq s \mid F\lceil_\rho \equiv 1 \ \wedge\ D_1\lceil_{\rho_1} \not\equiv 1]\ . \tag{2}$$

For a subset $\boldsymbol{Y}$ of T let $\boldsymbol{\mathcal{MT}_Y(C\lceil_\rho)}$ be the set of all conjunctive min-terms K_j in $\mathcal{MT}(C\lceil_\rho)$ such that $Y = \{i \in T \mid [B_i \in Z_i] \text{ is a test of } K_j\}$. Further define $\boldsymbol{\min(C\lceil_\rho)^Y}$ as the maximum number of variables which occur in a term of $\mathcal{MT}_Y(C\lceil_\rho)$.
Note that the property that $\mathcal{MT}_Y(C\lceil_\rho) \neq \emptyset$ for a restriction ρ and a set $Y \subseteq T$ implies that $\rho(B_i) = \rho_1(B_i) = \star$ for all $i \in Y$ – which will be denoted by $\rho_1(Y) = \star$. Using this notation we can bound probability 2 by

$$\begin{aligned}\sum_{Y\subset T,\ Y\neq\emptyset} \Big(\quad & \Pr[\rho_1(Y) = \star \mid F\lceil_\rho \equiv 1 \ \wedge\ D_1\lceil_{\rho_1} \not\equiv 1] \\ \cdot\ & \Pr[\min(C\lceil_\rho)^Y \geq s \mid F\lceil_\rho \equiv 1 \ \wedge\ D_1\lceil_{\rho_1} \not\equiv 1 \ \wedge\ \rho_1(Y) = \star]\ \Big)\ .\end{aligned} \tag{3}$$

The first factor of a single summand of this term can be estimated by

$$\Pr[\rho_1(Y) = \star \mid F\lceil_\rho \equiv 1 \ \wedge\ D_1\lceil_{\rho_1} \not\equiv 1] \ \leq\ \Pr[\rho_1(Y) = \star \mid D_1\lceil_{\rho_1} \not\equiv 1]\ . \tag{4}$$

From the precondition $D_1\lceil_{\rho_1} \not\equiv 1$ it follows directly that $\rho(B_i) \notin Z_i$ for each test $[B_i \in Z_i]$ of the disjunction D_1. Since the elements of ρ are chosen independently for all input variables we can bound probability 4 by

$$\Pr[\rho_1(Y) = \star \mid D_1\lceil_{\rho_1} \not\equiv 1] \ \leq\ (\Pr[\rho_1(B_1) = \star \mid D_1\lceil_{\rho_1} \not\equiv 1])^{|Y|} \ \leq\ \left(\tfrac{|\Sigma_I|\cdot p}{1+(|\Sigma_I|-1)\cdot p}\right)^{|Y|} \tag{5}$$

(remember that B_1 denotes a variable of disjunction D_1).
Let $\boldsymbol{\sigma} \in (\Sigma_I \cup \{\star\})^*$ be a satisfying assignment (resp. restriction) of a conjunctive min-term $K_j := \bigwedge_i [B_i \in Z_i]$ of $\mathcal{MT}_Y(C\lceil_\rho)$. Furthermore, let $\sigma(B_i) \in Z_i$ for all tests $[B_i \in Z_i]$ in K_j and $\sigma(B_i) = \star$ elsewhere. Similarly to ρ we can partition σ into σ_1 and σ_2 such that σ_1 gives the values of the variables in Y and the second part σ_2 the values of the variables B_i with $i \notin T$. From the definition of σ it follows that the set of variables B_i with $\sigma_2(B_i) \neq \star$ give a conjunctive min-term of $(C\lceil_\rho)\lceil_{\sigma_1}$. Hence σ decreases the number of disjunctions of $(C\lceil_\rho)\lceil_\sigma$ by one and the size bound of the conjunctive min-terms which has to be constructed by the induction to $s - |Y|$.
For such a set Y and an assignment σ_1 of the corresponding variables B_i define $\boldsymbol{\mathcal{MT}_Y^{\sigma_1}(C\lceil_\rho)} \subseteq \mathcal{MT}_Y(C\lceil_\rho)$ as the set of conjunctive min-terms K_j of $C\lceil_\rho$, where $K_j\lceil_{\sigma_1} \not\equiv 0$ and $\sigma_1(B_i) \in \Sigma_I$ for all variables B_i of

K_j with $i \in Y$. Furthermore, define $\mathbf{min}(\boldsymbol{C}\lceil_{\boldsymbol{\rho}})^{\boldsymbol{Y},\boldsymbol{\sigma_1}}$ as the maximum number of variables of a min-term in $\mathcal{MT}_Y^{\sigma_1}(C\lceil_\rho)$.
Let $\boldsymbol{\mathcal{W}}$ be the set of all assignments (resp. restrictions) $\sigma_1 \in (\Sigma_I \cup \{\star\})^{|Y|}$ of variables B_i in Y with $(D_1\lceil_{\rho_1})\lceil_{\sigma_1} \equiv 1$. Then the second factor of sum 3 can be bounded by

$$\begin{aligned}
&\Pr[\min(C\lceil_\rho)^Y \geq s \mid F\lceil_\rho \equiv 1 \ \wedge \ D_1\lceil_{\rho_1} \not\equiv 1 \ \wedge \ \rho_1(Y) = \star] \\
&\leq \ \sum_{\sigma_1 \in \mathcal{W}} \max_{\substack{\rho_1 \text{ with } \rho_1(Y)=\star \text{ and} \\ \rho_1(T\setminus Y) \in \Sigma_I^{|T|-|Y|}}} \Pr[\min((C\lceil_{\rho_1})\lceil_{\rho_2})^{Y,\sigma_1} \geq s \mid (F\lceil_{\rho_1})\lceil_{\rho_2} \equiv 1] \ .
\end{aligned}$$

To remove a variable B_i of Y with $\rho_1(B_i) = \sigma_1(B_i) = \star$ from C we will consider the conjunction $C\lceil_{B_i} := \bigwedge_{z \in \Sigma_I} C\lceil_{\star^{i-1} z \star^{n-i}}$ instead of C. Note that $C\lceil_{B_i}$ corresponds to C without variable B_i. For a nonempty subset S of variables of C define $\boldsymbol{C}\lceil_{\boldsymbol{S}} := (C\lceil_{B_i})\lceil_{S\setminus\{B_i\}}$ with $B_i \in S$. Further, let $C\lceil_\emptyset := C$.
For a restriction σ_1 of the variables B_i in Y define $S_{\sigma_1} := \{B_i \mid i \in Y \text{ and } \sigma_1[B_i] = \star\}$. From the definition of $\mathcal{MT}_Y^{\sigma_1}(C\lceil_\rho)$ it follows that K_j is a min-term of $\mathcal{MT}_Y^{\sigma_1}(C\lceil_\rho)$ if and only if K_j is also a min-term of $\mathcal{MT}_Y((C\lceil_{S_{\sigma_1}})\lceil_\rho)$. By induction we get

$$\Pr[\min(C\lceil_\rho)^Y \geq s \mid F\lceil_\rho \equiv 1 \ \wedge \ D_1\lceil_{\rho_1} \not\equiv 1 \ \wedge \ \rho_1(Y) = \star] \ \leq \ \sum_{\sigma_1 \in \mathcal{W}} \alpha_\rho^{s-|Y|} \ . \tag{6}$$

On the other hand it follows from the preconditions $D_1\lceil_{\rho_1} \not\equiv 1$ and $D_1\lceil_{\rho_1} \not\equiv 0$ that there are two assignments $\beta = \beta_1\beta_2$ and $\gamma = \gamma_1\gamma_2$ of the variables B_i of $C\lceil_\rho$ such that $(D_1\lceil_{\rho_1})\lceil_{\beta_1} \equiv 1$ and $(D_1\lceil_{\rho_1})\lceil_{\gamma_1} \equiv 0$. Hence, we can bound the number of restrictions $\sigma_1 \in \mathcal{W}$ by $|\Sigma_I|^{|Y|} - 1$. This implies the following bound:

$$\begin{aligned}
&\Pr[\min(C\lceil_\rho) \geq s \mid F\lceil_\rho \equiv 1 \ \wedge \ D_1\lceil_{\rho_1} \not\equiv 1] \\
&\leq \ \alpha_\rho^s \cdot \left(\left(1 + \frac{|\Sigma_I|^2 \cdot p}{1 + (|\Sigma_I| - 1) \cdot p} \cdot \frac{1}{\alpha_\rho}\right)^t - \left(1 + \frac{|\Sigma_I| \cdot p}{1 + (|\Sigma_I| - 1) \cdot p} \cdot \frac{1}{\alpha_\rho}\right)^t \right) \ = \ \alpha_\rho^s
\end{aligned} \tag{7}$$

∎

Proof of Lemma 3: To estimate the value of α_ρ we simplify the equation given in Lemma 2 as follows

$$\left(1 + |\Sigma_I| \cdot \frac{\beta}{\alpha_\rho}\right)^t = \left(1 + \frac{\beta}{\alpha_\rho}\right)^t + 1 \ . \tag{8}$$

Note that both terms can be approximated by

$$\left(1 + |\Sigma_I| \cdot \frac{\beta}{\alpha_\rho}\right)^t \ \leq \ e^{t \cdot |\Sigma_I| \cdot \frac{\beta}{\alpha_\rho}} \quad \text{and} \quad \left(1 + \frac{\beta}{\alpha_\rho}\right)^t + 1 \ \geq \ 2 \ .$$

Choosing $\alpha_\rho > \frac{t \cdot |\Sigma_I| \cdot \beta}{\ln 2}$ we get $(1 + |\Sigma_I| \cdot \beta/\alpha_\rho)^t < (1 + \beta/\alpha_\rho)^t + 1$. On the other hand, if $\alpha_\rho < |\Sigma_I| \cdot \beta$, $t \geq 2$ and $|\Sigma_I| \geq 3$ it holds:

$$\begin{array}{lrclcrcl}
 & |\Sigma_I| \cdot \beta & > & \alpha_\rho & \Longrightarrow & \frac{|\Sigma_I| - \sqrt[t]{2}}{\sqrt[t]{2} - 1} \cdot \beta & > & \alpha_\rho \\
\Longrightarrow & (|\Sigma_I| - \sqrt[t]{2}) \cdot \beta & > & (\sqrt[t]{2} - 1) \cdot \alpha_\rho & \Longrightarrow & 1 + |\Sigma_I| \cdot \frac{\beta}{\alpha_\rho} & > & \sqrt[t]{2} \cdot \left(1 + \frac{\beta}{\alpha_\rho}\right) \\
\Longrightarrow & \left(1 + |\Sigma_I| \cdot \frac{\beta}{\alpha_\rho}\right)^t & > & 2 \cdot \left(1 + \frac{\beta}{\alpha_\rho}\right)^t & \Longrightarrow & \left(1 + |\Sigma_I| \cdot \frac{\beta}{\alpha_\rho}\right)^t & > & \left(1 + \frac{\beta}{\alpha_\rho}\right)^t + 1 \ .
\end{array}$$

Choosing $\alpha_\rho \geq |\Sigma_I| \cdot \beta$ it follows that

$$\left(1 + |\Sigma_I| \cdot \frac{\beta}{\alpha_\rho}\right)^t \ > \ 2^{t \cdot |\Sigma_I| \cdot \frac{\beta}{\alpha_\rho}} \quad \text{and} \quad \left(1 + \frac{\beta}{\alpha_\rho}\right)^t + 1 \ < \ 2 \cdot e^{t \cdot \frac{\beta}{\alpha_\rho}} \ .$$

This implies that for any $\alpha_\rho \in \mathbb{R}$ with $|\Sigma_I| \cdot \beta \ \leq \alpha_\rho \leq \ t \cdot (|\Sigma_I| \cdot \ln 2 - 1) \cdot \beta / \ln 2$ it holds

$$\left(1 + |\Sigma_I| \cdot \frac{\beta}{\alpha_\rho}\right)^t > \left(1 + \frac{\beta}{\alpha_\rho}\right)^t + 1 \ .$$

Summarizing, equation 8 has its positive solutions for all p, t and Σ_I with $t \geq 2$ and $|\Sigma_I| \geq 3$ within the interval

$$\frac{t \cdot (|\Sigma_I| \cdot \ln 2 - 1) \cdot \beta}{\ln 2} \;<\; \alpha_\rho \;\leq\; \frac{t \cdot |\Sigma_I| \cdot \beta}{\ln 2} \;.$$

Since the terms of equation 8 are continuously over $\mathbb{R}^+$ there exists a positive solution for α_ρ. ∎

Proof of Lemma 4: The claim follows analogously to Lemma 2 if we allow arbitrary conjunctions for $C\lceil_{\rho_{[p,\Sigma_I]}}$ (and not only conjunctive min-terms). Hence, we can choose Y as the set of all variables B_i of T with $\rho(B_i) = \star$. This implies that $\mathcal{W}$ will be chosen as the set of all assignments σ_1 of the variables given by Y with $(D_1\lceil_{\rho_{[p,\Sigma_I]}})\lceil_{\sigma_1} \equiv 1$. Note that the size of $\mathcal{W}$ is still bounded by $|\mathcal{W}| \leq |\Sigma_I|^{|Y|} - 1$. ∎

Proof of Theorem 3: To give an upper bound for $\mathcal{E}(f,d,s,n)$ we will use a probabilistic restriction $\rho_{[p,\Sigma_I]}$, noted by ρ for short.
We restrict ourself to circuits C of $\boldsymbol{\ell}_{\mathbf{out}} = |\Sigma|$ output gates $v_1, \ldots, v_{\ell_{\text{out}}}$ where C outputs $q_i \in \Sigma$ on input $w \in \Sigma_I^*$ iff $\text{res}_{v_i}(w) = 1$. Further, let C_i be the subcircuit of C with one output gate v_i and define $\Sigma := \{q_1, \ldots, q_{\ell_{\text{out}}}\}$.

Lemma 5 *Let* $f \in \mathcal{F}^{\text{last}}_{\text{o-diff}}$ *then it holds* $\mathcal{E}(f,2,s,n) \leq \frac{|\Sigma_I|^2 - 1}{|\Sigma_I|^2} + |\Sigma_I|^{-\Omega(n/s)}$.

Proof: Let C be an ub-circuit of depth 2. From the output restrictions of C we can conclude that C computes f correctly on an input $w \in \Sigma_I^n$ only if there exists exactly one output gate v of C with $\text{res}_v(w) = 1$. From Definition 8 it follows that the output rank of each function $f \in \mathcal{F}^{\text{last}}_{\text{o-diff}}$ is given by Σ. Hence, each circuit C computing a function $f \in \mathcal{F}^{\text{last}}_{\text{o-diff}}$ correctly has no constant output gate.
In the following we focus on a conjunction of disjunctions. Note, that the analysis of a disjunction of conjunctions follows analogously.

Figure 3: A depth 2 ub-circuit.

The proof of Lemma 5 follows by induction over the number β of disjunctions with indegree at least d with $d \ll n$. For $\beta = 0$ it holds:

Lemma 6 *Let* $f \in \mathcal{F}^{\text{last}}_{\text{o-diff}}$ *and* C *be a ub-circuit of depth 2 where all gates of depth 1 have maximum indegree* d*. Then the fraction of all inputs of length* n *where* C *computes* f *correctly is at most* $\frac{|\Sigma_I|^2-1}{|\Sigma_I|^2} + 2^{-\Omega(n/d)}$.

Proof: From Corollary 4 it follows that for a probabilistic restriction $\rho := \rho_{[p,\Sigma_I]}$ the probability that $C_i\lceil_\rho$ can be described by a disjunction of conjunctions $\bigvee_j K_{i,j}$ of maximal indegree s is at least $1 - \alpha_\rho^s$. This also holds, if all sets $\text{accept}(K_{i,j})$ are pairwise disjoint.
Let $\boldsymbol{D_i := \bigvee_{j \leq d} K_{i,j}}$ be the disjunction computing $C_i\lceil_\rho$, then

$$\text{accept}(D_i) \;=\; \bigcup_{j \leq d} \text{accept}(K_{i,j}) \;.$$

Define $\boldsymbol{\text{accept}(f,i)}$ as the set of all inputs $w \in \Sigma_I^n$ with $f(w) = q_i$. Since f is strictly diffluent we can conclude from Theorem 2 and the size bound of the conjunctions that for all values of i and j

$$\frac{|\text{accept}(K_{i,j}) \setminus \text{accept}(f\lceil_\rho, i)|}{|\text{accept}(K_{i,j})|} \;\geq\; \frac{1}{|\Sigma_I|} \;.$$

Furthermore, if the number of input variables of $f\lceil_\rho$ is greater than s it holds

$$\begin{aligned}|\mathrm{accept}(D_i)\setminus \mathrm{accept}(f,i)| &= \sum_j |\mathrm{accept}(K_{i,j})\setminus \mathrm{accept}(f\lceil_\rho,i)| \\ &\geq \frac{1}{|\Sigma_I|}\sum_j |\mathrm{accept}(K_{i,j})| \;=\; \frac{|\mathrm{accept}(D_i)|}{|\Sigma_I|} \;.\end{aligned}$$

Hence, $C\lceil_\rho$ computes f for a fraction of at most $1-\frac{1}{|\Sigma_I|}$ of all inputs w with $C_i\lceil_\rho(w)=1$ correctly.
Let $\boldsymbol{\mathrm{fault}(C_i\lceil_\rho, f, i)}$ be the set of all inputs w where C_i generates an error, i.e. $C_i\lceil_\rho(w)\neq f\lceil_\rho(w)$, then

$$\mathrm{fault}(C_i\lceil_\rho,f,i) \;=\; (\mathrm{accept}(C_i\lceil_\rho)\setminus \mathrm{accept}(f\lceil_\rho,i)) \cup (\mathrm{accept}(f\lceil_\rho,i)\setminus \mathrm{accept}(C_i\lceil_\rho)) \;.$$

Note that for each function $f\in\mathcal{F}^{\mathrm{last}}_{\circ-\mathrm{diff}}$ there exists at least one element $q_i\in\Sigma$ such that for at least $|\Sigma_I|^{n-1}$ input w of length n holds $f(w)=q_i$. Therefore, we get

$$\begin{aligned}\frac{|\mathrm{fault}(C_i\lceil_\rho,f,i)|}{|\Sigma_I^n|} &\geq \frac{|\mathrm{fault}(C_i\lceil_\rho,f,i)|}{|\Sigma_I|\cdot|\mathrm{accept}(f\lceil_\rho,i)|} \\ &\geq \frac{|\mathrm{accept}(C_i\lceil_\rho)\setminus \mathrm{accept}(f\lceil_\rho,i)|}{|\Sigma_I|\cdot|\mathrm{accept}(C_i\lceil_\rho)\cap \mathrm{accept}(f\lceil_\rho,i)|} \;\geq\; \frac{1}{|\Sigma_I|^2}\;.\end{aligned}$$

In the following we will investigate the probability that $\#\mathrm{Var}(C_i\lceil_\rho)>s$ where $\boldsymbol{\#Var(C)}$ denotes the number of variables of C. Let $p=\frac{1}{k\cdot|\Sigma_I|^2\cdot d}$ for a constant $k\geq 2$. Then the probability that $C_i\lceil_\rho$ has a conjunction of more than $\frac{n}{2\cdot k\cdot c\cdot|\Sigma_I|^2\cdot d}$ variables for a constant c is at most $k^{-\Omega(n/d)}$. On the other hand the probability that $\#\mathrm{Var}(C_i\lceil_\rho)\leq\frac{n}{2\cdot k\cdot c\cdot|\Sigma_I|^2\cdot d}$ can be bounded by:

$$\begin{aligned}\Pr[\#\mathrm{Var}(C_i\lceil_\rho)=\frac{p\cdot n}{c}] &\leq 2\cdot\frac{(p\cdot n)^{p\cdot n/c}}{(p\cdot n/c)!}\cdot e^{-p\cdot n} \\ &\leq 2\cdot\frac{1}{\sqrt{2\cdot\pi\cdot p\cdot n/c}}\cdot\left(\frac{c\cdot e}{p\cdot n}\right)^{p\cdot n/c}\cdot\left(\frac{(p\cdot n)^{1/c}}{e}\right)^{p\cdot n} \\ &\leq \sqrt{\frac{c}{p\cdot n}}\cdot\left(\frac{(c\cdot e)^{1/c}}{e}\right)^{p\cdot n}\end{aligned}$$

This implies that $\displaystyle\sum_{j=1}^{p\cdot n/c}\Pr[\#\mathrm{Var}(C_i\lceil_\rho)=j]\leq\frac{p\cdot n}{c}\cdot\Pr[\#\mathrm{Var}(C_i\lceil_\rho)=\frac{p\cdot n}{c}]\leq\sqrt{\frac{p\cdot n}{c}}\cdot\left(\frac{(c\cdot e)^{1/c}}{e}\right)^{p\cdot n}$.

For $c>1$ we get $\Pr[\#\mathrm{Var}(C_i\lceil_\rho)\leq\frac{p\cdot n}{c}]\in 2^{-\Omega(p\cdot n)}=2^{-\Omega(n/d)}$. Hence, the probability that $C_i\lceil_\rho$ computes f with an error fraction greater then $1/|\Sigma_I|^2$ is at least $1-2^{-\Omega(n/d)}$.
Since the set $\mathrm{accept}(C_i\lceil_\rho)$ are not necessarily disjoint we have to investigate the circuits $C_i\lceil_\rho$ more meticulously to get a meaningful error fraction of C_i. In a second step we will now transfer the error fraction of $C_i\lceil_\rho$ to C_i.
For a disjunction of conjunctions $C_i\lceil_\rho$ define $\boldsymbol{\mathrm{MaxIn}(C_i\lceil_\rho)}$ as the maximum indegree of a conjunction of $C_i\lceil_\rho$.
Further, let $\boldsymbol{\mathcal{C}_1}$ be the set of all conjunctions $C_i\lceil_\sigma$ with $\sigma\in(\Sigma_I\cup\{\star\})^n$ such that $\#\mathrm{Var}(C_i\lceil_\sigma)>\mathrm{MaxIn}(C_i\lceil_\sigma)$ and define for $j,m\in\mathbb{N}$:

$$\begin{aligned}\boldsymbol{\mathcal{C}_1^{[j]}} &:= \{C_i\lceil_\sigma\in\mathcal{C}_1 \mid \#\mathrm{Var}(C_i\lceil_\sigma)=j\} \\ \boldsymbol{\mathcal{C}_1^{\langle m\rangle}} &:= \{C_i\lceil_\sigma\in\mathcal{C}_1 \mid \mathrm{MaxIn}(C_i\lceil_\sigma)\geq m\} \\ \boldsymbol{\mathcal{C}_1^{[j],\langle m\rangle}} &:= \{C_i\lceil_\sigma\in\mathcal{C}_1 \mid \#\mathrm{Var}(C_i\lceil_\sigma)=j \text{ and } \mathrm{MaxIn}(C_i\lceil_\sigma)\geq m\}\;.\end{aligned}$$

Note that the error bound given above holds for all circuits $C_i\lceil_\sigma\in\mathcal{C}_1$. Now define $\boldsymbol{\mathcal{C}_2}$ as the set of all circuits $C_i\lceil_\sigma$ with $\#\mathrm{Var}(C_i\lceil_\sigma)=\mathrm{MaxIn}(C_i\lceil_\sigma)$ and

$$\boldsymbol{\mathcal{C}_2^{[m]}} \;:=\; \{C_i\lceil_\sigma\in\mathcal{C}_2 \mid \#\mathrm{Var}(C_i\lceil_\sigma)=m\}\;.$$

It follows from the probabilistic restriction and the estimation given above that for $\boldsymbol{\ell}:=\frac{n}{2\cdot k\cdot c\cdot|\Sigma_I|^2\cdot d}$ it holds

$$\frac{|\mathcal{C}_2|}{|\mathcal{C}_1\cup\mathcal{C}_2|}\leq\frac{|\mathcal{C}_2\cup\bigcup_{i=1}^{\ell-1}\mathcal{C}_1^{[i]}|}{|\mathcal{C}_1\cup\mathcal{C}_2|}\in 2^{-\Omega(n/d)}\;.$$

Hence, there exists at least one $m \geq \ell$ such that

$$\frac{|\mathcal{C}_2^{[m]}|}{|\mathcal{C}_1^{[m]} \cup \mathcal{C}_2^{[m]}|} \in 2^{-\Omega(n/d)} \ .$$

Choose m_0 minimal such that m_0 fulfils this condition.
In the next step we partition the set $\mathcal{C}_1^{[m_0]} \cup \mathcal{C}_2^{[m_0]}$ over the set of unrestricted variables $\mathcal{V}$ into the subsets $\mathcal{C}_1^{[m_0],\mathcal{V}} \cup \mathcal{C}_2^{[m_0],\mathcal{V}}$ with $\mathcal{C}_j^{[m_0],\mathcal{V}} \in \mathcal{C}_j^{[m_0]}$ for $j \in \{1,2\}$. Then for each variable set $\mathcal{V}$ the union of the domains of the conjunctions in $\mathcal{C}_1^{[m_0],\mathcal{V}} \cup \mathcal{C}_2^{[m_0],\mathcal{V}}$ expanded by the fixed variables of ρ is an exact cover of Σ_I^n. Therefore, there exists a set of variables $\mathcal{V}_0$ such that:

$$\frac{|\mathcal{C}_2^{[m_0],\mathcal{V}_0}|}{|\mathcal{C}_1^{[m_0],\mathcal{V}_0} \cup \mathcal{C}_2^{[m_0],\mathcal{V}_0}|} \in 2^{-\Omega(n/d)} \ .$$

Let $\boldsymbol{\mathcal{A}(C_i, f, i)}$ be the number of inputs w on which C_i decides whether $f(w) = q_i$ correctly. Assuming that the circuits in $\mathcal{C}_2^{[m_0],\mathcal{V}_0}$ computes always the correct result, we can bound $\mathcal{A}(C_i, f, i)$ as follows

$$\begin{aligned}
\mathcal{A}(C_i, f, i) &\leq 1 - \frac{1}{|\Sigma_I|^2} + \frac{|\Sigma_I|^{m_0} \cdot |\mathcal{C}_2^{[m_0],\mathcal{V}_0}|}{|\Sigma_I|^n} \\
&\leq 1 - \frac{1}{|\Sigma_I|^2} + \frac{|\Sigma_I|^{m_0} \cdot |\Sigma_I|^{n-m_0} \cdot 2^{-\Omega(n/d)}}{|\Sigma_I|^n} \\
&= 1 - \frac{1}{|\Sigma_I|^2} + 2^{-\Omega(n/d)} \ .
\end{aligned}$$

∎

Let $c \cdot \log_2 |\Sigma_I|$ be the constant implicitly given by Lemma 6. In the following we will estimate the error fraction of a depth 2 circuit C with $\beta > 0$.

Lemma 7 *Let $f \in \mathcal{F}_{\circ-\mathrm{diff}}^{\mathrm{last}}$ and C be a ub-circuit of depth 2 with $\beta > 0$ gates of depth 1 which have an indegree greater than d. Then the fraction of inputs of length n where C computes f correctly is at most*

$$\frac{|\Sigma_I|^2 - 1}{|\Sigma_I|^2} + (\beta + 1)^{\ell \cdot n/d^2} \cdot |\Sigma_I|^{-c \cdot n/d}$$

with $\ell := 2 \cdot (|\Sigma_I|^2 + 1) \cdot c \cdot |\Sigma_I| \cdot \log_2 |\Sigma_I|$.

Proof: The proof follows by induction over the number of variables n. Note that the claim follows directly for $n = 0$ from the chosen fraction and for $\beta = 0$ from Lemma 6. Now, assume that the claim holds for all $n \leq n_0$ and $\beta \leq \beta_0$.
Again, let C_i be a conjunction von disjunctions with $n = n_0 + 1$ variables and $\beta > \beta_0$ disjunctions with indegree at least $d + 1$. Then there exists a variable B_j that supplies at least $d \cdot \beta / n$ of these disjunctions. Define $\boldsymbol{\beta_{B_j = x}}$ with $x \in \Sigma_I$ as be number of disjunctions of $C_i \lceil_{\star^{j-1} x \star^{n-j}}$ with indegree greater then d. By induction it follows that:

$$\begin{aligned}
\mathcal{A}(C_i, f, i) &= \frac{1}{|\Sigma_I|} \cdot \sum_{x \in \Sigma_I} \mathcal{A}(C_i \lceil_{\star^{j-1} x \star^{n-j}}, f, i) \\
&\leq \frac{1}{|\Sigma_I|} \cdot \sum_{x \in \Sigma_I} \left(\frac{|\Sigma_I|^2 - 1}{|\Sigma_I|^2} + (\beta_{B_j = x} + 1)^{\ell \cdot (n-1)/d^2} \cdot |\Sigma_I|^{-c \cdot (n-1)/d} \right) \\
&= \frac{|\Sigma_I|^2 - 1}{|\Sigma_I|^2} + \frac{1}{|\Sigma_I|} \cdot \sum_{x \in \Sigma_I} (\beta_{B_j = x} + 1)^{\ell \cdot (n-1)/d^2} \cdot |\Sigma_I|^{-c \cdot (n-1)/d} \\
&\leq \frac{|\Sigma_I|^2 - 1}{|\Sigma_I|^2} + (\beta + 1)^{\ell \cdot n/d^2} \cdot |\Sigma_I|^{c/d - 1 - c \cdot n/d} \cdot \sum_{x \in \Sigma_I} \left(\frac{\beta_{B_j = x} + 1}{\beta + 1} \right)^{\ell \cdot n/d^2}
\end{aligned}$$

Since B_j supplies at least $d \cdot \beta / n$ disjunctions there exists an input symbol $x \in \Sigma_I$ that satisfies at least $\frac{d \cdot \beta}{n \cdot |\Sigma_I|}$ of the testing gates $[B_j \in Z_k]$. Hence, $\beta_{B_j = x} < \beta \cdot (1 - d/(|\Sigma_I| \cdot n))$. This implies the following estimate

$$\sum_{x \in \Sigma_I} \left(\frac{\beta_{B_j = x} + 1}{\beta + 1} \right)^{\ell \cdot n/d^2} \leq |\Sigma_I| - 1 + \left(\frac{\beta \cdot (1 - d/(|\Sigma_I| \cdot n) + 1}{\beta + 1} \right)^{\ell \cdot n/d^2}$$

$$\begin{aligned} &= |\Sigma_I| - 1 + \left(1 - \frac{d \cdot \beta}{|\Sigma_I| \cdot n \cdot (\beta+1)}\right)^{\ell \cdot n/d^2} \\ &\leq |\Sigma_I| - 1 + \exp\left(\frac{-\ell}{2 \cdot d \cdot |\Sigma_I|}\right) \\ &\leq |\Sigma_I| - 1 + |\Sigma_I|^{-(|\Sigma_I|^2+1)\cdot c/d} \end{aligned}$$

for a sufficient large n. Finally, we have to show that for some adequate d it holds

$$|\Sigma_I|^{c/d-1} \cdot \left(|\Sigma_I| - 1 + |\Sigma_I|^{-(|\Sigma_I|^2+1)\cdot c/d}\right) \quad \leq \quad 1 \; . \tag{9}$$

Considering the function $g(d) := |\Sigma_I|^{c/d-1} \cdot \left(|\Sigma_I| - 1 + |\Sigma_I|^{-k \cdot c/d}\right)$ we get

$$g'(d) = \frac{c \cdot |\Sigma_I|^{c/d-1} \cdot \ln(|\Sigma_I|)}{d^2} \cdot \Big((k-1) \cdot |\Sigma_I|^{-k\cdot c/d} - (|\Sigma_I| - 1)\Big) \; .$$

Since $g'(d) > 0$ for all $k > 2 \cdot |\Sigma_I|$ and $d > k \cdot c / \log_{|\Sigma_I|} 2$ the function $g(d)$ is strictly monotone growing. On the other hand $g(d)$ tends towards 1 in the limit. Hence, inequality 9 holds for all $d > k \cdot c/\log_{|\Sigma_I|} 2$ with $k := |\Sigma_I|^2 + 1$.

∎

Since β is always bounded by the size of the circuits – i.e. $\beta \leq |\Sigma_I|^s$ – the claim of Lemma 5 follows directly by choosing

$$d \;\geq\; \begin{cases} \ell \cdot s & \text{if } c \geq 1 \;, \\ 2 \cdot (|\Sigma_I|^2+1) \cdot |\Sigma_I| \cdot \log_2 |\Sigma_I| \cdot s & \text{elsewhere.} \end{cases}$$

∎

Within the next lemma we will generalise Lemma 5 to t–depth bounded circuits.

Lemma 8 *Let $f \in \mathcal{F}^{\text{last}}_{\circ-\text{diff}}$ then it holds:*

$$\mathcal{E}(f,t,s,n) \leq \frac{|\Sigma_I|^2-1}{|\Sigma_I|^2} + \begin{cases} |\Sigma_I|^{-\Omega(n/s^{t-1})} & \text{if } s > \sqrt[t]{n} \\ |\Sigma_I|^{-\Omega((n/s)^{1/t})} & \text{if } s < \sqrt[t]{n} \;. \end{cases}$$

Proof: Analogously to Lemma 5 we can divide the proof of Lemma 8 into two parts.
As in the case of depth 2 circuit we restrict ourself to circuits with ℓ_{out} output gates $v_1, \ldots, v_{\ell_{\text{out}}}$. Further, we assume that all paths from an input to an output gate have the same length and alternating types of gates. Hence, all gates of depth $k > 0$ are only connected to gates of depth $k-1$. We assume that all gates of odd depth are disjunctions and all gates of even depth conjunctions (excepting the testing gates). The alternative case follows analogously. Remember that the input gates are testing gates $[x_j \in Z_j]$, which have depth 0.

Lemma 9 *Let $f \in \mathcal{F}^{\text{last}}_{\circ-\text{diff}}$ and C be a circuit of size $|\Sigma_I|^s$ and depth t where all the gates of depth 1 have maximum indegree $d \geq s$. Then the fraction of all inputs of length n where C computes f correctly is at most*

$$\frac{|\Sigma_I|^2-1}{|\Sigma_I|^2} + \begin{cases} |\Sigma_I|^{-\Omega(n/d^{t-1})} & \text{if } d > \sqrt[t]{n}, \\ |\Sigma_I|^{-\Omega((n/d)^{1/(t-1)})} & \text{if } d < \sqrt[t]{n} \;. \end{cases}$$

Proof: The proof follows by induction over the depth t of the subcircuits C_i. Note that for $t = 2$ the claim follows directly from Lemma 6.
For $t > 2$ let $r_1 := \max\{d, (n/d)^{1/(t-1)}\}$ and $p_1 := \frac{1}{k \cdot |\Sigma_I|^2 \cdot d}$ with $k > 2/\ln 2$. Note that the number of gates of depth 2 is bounded by $|\Sigma_I|^s \leq |\Sigma_I|^d$. So it follows from Lemma 2 that for $\rho := \rho_{[p,\Sigma_I]}$ with probability at least $(1 - |\Sigma_I|^s \cdot \alpha_\rho^{r_1}) \in (1 - 2^{-\Omega(r_1)})$ we can describe $C_i\lceil_\rho$ by a $t-1$-depth and $|\Sigma_I|^s$-size bounded circuit where all gates of depth 1 has a maximum indegree of r_1. Note that $\alpha_\rho < 1/2$ (see Lemma 3).
As it is shown in Lemma 6 we can bound the probability that $\#\text{Var}(C_i\lceil_\rho)$ is greater then $p_1 \cdot n/c$ for a constant $c > 1$ by $1 - 2^{-\Omega(p_1 \cdot n)}$.

Note that $r_1 = d$ if and only if $d \geq \sqrt[t]{n}$. Assume that $f_{C_i}\lceil_\rho$ fulfils the condition given above. Then it follows by induction that the error fraction of $f_{C_i}\lceil_\rho$ is at least

$$1 - \frac{|\Sigma_I|^2 - 1}{|\Sigma_I|^2} + |\Sigma_I|^{-\Omega(p_1 \cdot n/(c \cdot r_1{}^{t-2}))} .$$

where $p_2 := \frac{1}{k \cdot |\Sigma_I|^2 \cdot r_1}$ and $r_2 := \max\{r_1, (p_1 \cdot n/(c \cdot r_1))^{1/(t-2)}\} = r_1$. Analogously to Lemma 6 we can bound the fraction of all inputs w where C_i decides whether $f(w) = q_i$ correctly by

$$\leq \frac{|\Sigma_I|^2 - 1}{|\Sigma_I|^2} + |\Sigma_I|^{-\Omega(p_1 \cdot n/(c \cdot r_1{}^{t-2}))} + 2^{-\Omega(p_1 \cdot n)} + 2^{-\Omega(r_1)} .$$

The claim follows directly by the two alternative choices of r_1. ∎

Analogously to Lemma 7 we can expand Lemma 9 to circuits without indegree restrictions. Let $c \cdot \log_2 |\Sigma_I|$ be the constant implicitly given in Lemma 9. Further, let β be the number of gates of depth 1 with indegree greater that d where $d \ll n$.

Lemma 10 *Let $f \in \mathcal{F}^{\text{last}}_{\circ-\text{diff}}$ and C be a circuit of size $|\Sigma_I|^s$ and depth t where the number of gates of depth 1 with indegree greater than $d \geq \max\{s, \sqrt[t]{n} + 1\}$ is given by β. Then the fraction of inputs of length n where C computes f correctly is at most*

$$\frac{|\Sigma_I|^2 - 1}{|\Sigma_I|^2} + (\beta + 1)^{\ell/d \ \cdot n/d^{t-1}} \cdot |\Sigma_I|^{-\Omega(n/d^{t-1})}$$

where $\ell := 2 \cdot (|\Sigma_I|^2 + 1) \cdot c \cdot |\Sigma_I| \cdot \log_2 |\Sigma_I|$.

Proof: The proof follows analogously to Lemma 7. Therefore we will only present the most important inequalities of the proof. Again, the proof is done by induction over the number of variables n. Note that for $n = 0$ the claim follows directly from the chosen error bound. Furthermore, the claim follows from Lemma 9 if $\beta = 0$. In the following we assume that the claim of Lemma 10 holds for all $n \leq n_0$ and $\beta \leq \beta_0$.

Let C_i be an adequate circuit with $n_0 + 1$ input gates and $\beta > \beta_0$ gates of depth 1 with indegree at least $d+1$. Hence, there exists an input variable B_j that supplies at least $d \cdot \beta/n$ of these gates. Note that:

$$\mathcal{A}(C_i, f, i) \ = \ \frac{1}{|\Sigma_I|} \cdot \sum_{z \in \Sigma_I} \mathcal{A}(C_i\lceil_{\star^{j-1} z \star^{n-j}}, f, i) \ .$$

Analogously to Lemma 7 we can estimate the number $\mathcal{A}(C_i, f, i)$ of inputs where C_i computes f correctly as follows

$$\begin{aligned}\mathcal{A}(C_i, f, i) \leq \frac{|\Sigma_I|^2 - 1}{|\Sigma_I|^2} + \ & (\beta + 1)^{\ell/d \ \cdot n/d^{t-1}} \cdot |\Sigma_I|^{-c \cdot n/d^{t-1}} \\ & \cdot \left(|\Sigma_I|^{c/d^{t-1} - 1} \cdot (|\Sigma_I| - 1 + |\Sigma_I|^{-(|\Sigma_I|^2+1) \cdot c/d^{t-1}}) \right) \ .\end{aligned}$$

Let $g(d) := |\Sigma_I|^{c/d^{t-1} - 1} \cdot (|\Sigma_I| - 1 + |\Sigma_I|^{-(|\Sigma_I|^2+1) \cdot c/d^{t-1}})$. Since

$$g'(d) = \frac{c \cdot (t-1) \cdot \ln(|\Sigma_I|)}{|\Sigma_I| \cdot d^t} \cdot h(d)$$

with $h(d) := (|\Sigma_I|^2 + 1 - |\Sigma_I|^{-c/d^{t-1}}) \cdot |\Sigma_I|^{-c \cdot |\Sigma_I|^2/d^{t-1}} - |\Sigma_I| + 1$ the function $g(d)$ is strictly monotone growing for all $d^{t-1} \geq c \cdot |\Sigma_I|^2$. On the other hand $\lim_{d \to \infty} g(d) = 1$. Hence, $g(d) < 1$ for all $d \geq \sqrt[t-1]{c \cdot |\Sigma_I|^2}$. The claim follows directly. ∎

Note that for $d < \sqrt[t]{n - 1}$ the value of $\mathcal{A}(C_i, f, i)$ is bounded by

$$\begin{aligned}\mathcal{A}(C_i, f, i) \leq \frac{|\Sigma_I|^2 - 1}{|\Sigma_I|^2} + \ & (\beta + 1)^{\ell/d \ \cdot n/d^{t-1}} \cdot |\Sigma_I|^{-c \cdot n/d^{t-1}} \\ & \cdot \left(|\Sigma_I|^{c/d^{1/(t-1)} - 1} \cdot \sum_{z \in \Sigma_I} \left(\frac{\beta_{B_j = z} + 1}{\beta + 1} \right)^{\ell/d \ \cdot (n/d)^{1/(t-1)}} \right) \ ,\end{aligned}$$

where the sum of the inequality can only be estimated by

$$|\Sigma_I| - 1 + |\Sigma_I|^{-(|\Sigma_I|^2+1) \cdot c/d \ \cdot n^{-\frac{t-2}{t-1}}}$$

using the technique given in Lemma 7. However, this estimation does not seams to give an sufficient bound for the range of correctness. On the other hand, this problem can easily be solved by considering a new circuit C' of depth $t+1$ instead of C, that is generated by adding constant indegree nodes of depth 1 to any input gate of C. So it follows from Lemma 9:

Lemma 11 *Let $f \in \mathcal{F}^{\mathrm{last}}_{\circ-\mathrm{diff}}$ and C be a ub–circuit of size $|\Sigma_I|^s$ and depth t. Further assume that C has at least one gate of depth 1 with indegree $d \geq s$ and $d < \sqrt[t]{n}$. Then the fraction of all inputs of length n where C computes f correctly is at most*

$$\frac{|\Sigma_I|^2 - 1}{|\Sigma_I|^2} + \begin{cases} |\Sigma_I|^{-\Omega(n/d^t)} & \text{if } d > \sqrt[t+1]{n} \\ |\Sigma_I|^{-\Omega((n/d)^{1/t})} & \text{if } d < \sqrt[t+1]{n}\ . \end{cases}$$

Choosing $d := s$ Lemma 8 follows form the lemmata shown above and the fact that $\beta \leq |\Sigma_I|^s$. ∎

The claim of Theorem 3 follows directly by combining the results of Lemma 5 and 8. ∎