A note on the depth of optimal fanout-bounded prefix circuits

I. S. Sergeev*

Abstract

It is shown that the minimal depth of an optimal prefix circuit (i.e., a zero-deficiency circuit) on N inputs with fanout bounded by k is $\log_{\alpha_k} N \pm O(1)$, where α_k is the unique positive root of the polynomial $2 + x + x^2 + \ldots + x^{k-2} - x^k$. This bound was previously known in the cases $k = 2$ and $k = \infty$.

Introduction

Let $(\mathbb{S}, \circ)$ be a semigroup. The set of functions

$$s_i = x_1 \circ x_2 \circ \ldots \circ x_i, \qquad 1 \le i \le N, \tag{1}$$

is called the system of *prefix sums* of variables $x_1, \ldots, x_N$ taking values in $\mathbb{S}$. Circuits of functional elements over the basis $\{x \circ y,\ x\}$ that implement the system (1) are called *prefix circuits*. The number N (of circuit inputs) is called the *width* of a circuit. By the *complexity* of a circuit we will (as usual) mean the total number of binary elements "$\circ$" in it. The need for identity elements appears only when the circuit fanout is bounded. The *depth* of a circuit is the maximum number of elements (of both types) in an input-output path.

We consider *universal* prefix circuits that correctly compute sums regardless of the choice of a semigroup $\mathbb{S}$. It is easy to verify that in a minimal (i.e., not containing elements unconnected to outputs) universal circuit, only interval sums are computed via operations of the form $p_1 \circ p_2$, where $p_1 = x_i \circ x_{i+1} \circ \ldots \circ x_j$ and $p_2 = x_{j+1} \circ x_{j+2} \circ \ldots \circ x_l$. If a node in the circuit computes the sum $x_i \circ \ldots \circ x_j$, then j is called the *index* of this node.

Obviously, all sums s_i can be computed sequentially, with a minimum number of $N - 1$ operations "$\circ$". The complexity C and the depth D of a prefix circuit of width N are related as $C + D \ge 2N - 2$ [1, 4], so the complexity of parallel prefix circuits cannot be significantly less than $2N$. Prefix circuits for which the equality $C + D = 2N - 2$ holds are called *optimal* or circuits with zero deficiency.

As is known, an optimal prefix circuit of depth D on N inputs (when it exists) has the following structure. Its elements either belong to the *framework*

*e-mail: isserg@gmail.com

tree of depth D, which is the subcircuit computing the sum of all inputs, or are outputs of the circuit. Each of these two sets of elements has cardinality $N-1$, but D elements of the *principal chain*, i.e., the chain connecting the first input with the last output of the circuit, belong to both sets[1]. Therefore, a circuit has complexity $2N-D-2$. For more details, see, e.g., [5].

Let $D(N,k)$ denote the minimum possible depth of an optimal circuit on N inputs with fanout bounded by k.

It is shown in [5] that

$$D(N,\infty) = d = \log_\varphi -O(1) \approx 1.44\log_2 N - O(1), \tag{2}$$

where Φ_{d+3} is the nearest number from the Fibonacci sequence $\{\Phi_m\}$ to $N+1$ from above, and $\varphi = \frac{1+\sqrt{5}}{2}$. From [2, 3] it follows that $D(N,2) = \lfloor \log_2 N \rfloor + \lfloor \log_2(2N/3) \rfloor$. Exact or at least asymptotic closed-form estimates for $D(N,k)$, where $2 < k < \infty$, have apparently not yet been obtained, despite the fact that, for example, in [3] optimal fanout-bounded circuits of extreme width were constructed.

Let α_k denote the unique positive root of the polynomial $P_k(x) = 2 + x + x^2 + \ldots + x^{k-2} - x^k$. Further, we will prove

Theorem. *$D(N,k) = \log_{\alpha_k} N \pm O(1)$.*

It is easy to verify that $\alpha_k \to \frac{1+\sqrt{5}}{2}$ as $k \to \infty$, which is consistent with (2). In particular, the theorem implies $D(N,3) \sim 1.65\ldots \cdot \log_2 N$, $D(N,4) \sim 1.54\ldots \cdot \log_2 N$ and already $D(N,9) \lesssim 1.45 \log_2 N$.

Proof of the theorem

Consider an optimal prefix circuit of depth D with N inputs. Let its principal chain be formed by a sequence of nodes $v_0, v_1, \ldots, v_D$, where v_0 coincides with input x_1 and an arbitrary node v_d is located at depth d.

The nodes of the principal chain naturally partition the circuit into segments. If the sums s_t and s_{t+w}, respectively, are calculated at nodes v_d and v_{d+1}, then the d-th segment includes the inputs and nodes with indices in the interval $[t,\, t+w]$. The parameter w denotes the segment width. The structure of a segment of an optimal circuit is shown in Fig. 1 (the notation is standard, see, e.g., [3, 5]). There $h = D - d$.

The segment's construction is determined by two trees-subcircuits: a binary tree directed from the inputs $x_{t+1}, \ldots, x_{t+w}$ to the root u_d, and a tree consistent with it, directed from the root v_d to the outputs $s_t, \ldots, s_{t+w-1}$. The fanout of the second tree is bounded by k. Tree consistency means that the second tree employs exactly the interval sums calculated by the first tree. In particular, all descendant neighbors of the node v_d receive second inputs strictly from nodes in the chain connecting x_{t+1} and u_d.

[1]Moreover, upon transposition, i.e., reversing the direction of the circuit, both sets are mapped into each other.

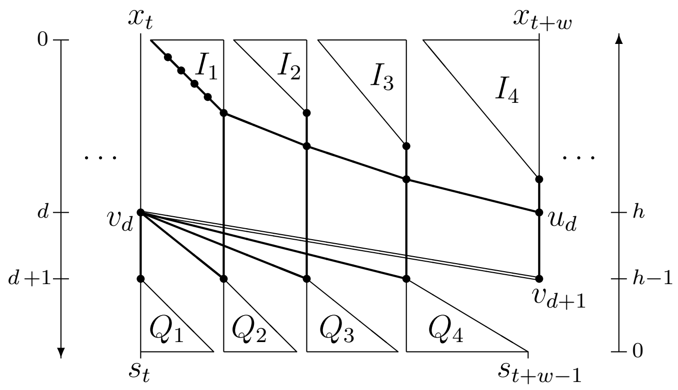

Figure 1: Structure of a segment of an optimal prefix circuit

The structure of one segment is independent of the structure of the other segments. Therefore, the maximum width of a circuit of a given depth and fanout is the sum of the maximum possible widths of the segments.

Let $w_k(d,h)$ denote the maximum width of a pair of consistent trees, the first of which has depth $\le d$ and the second has depth $\le h$ and fanout bounded by k. By $w_k(D)$ we denote the maximum width of an optimal depth-D fanout-k circuit. We introduce the notation

$$w_k^*(D) = \sum_{d=0}^{D} w_k(d, D-d).$$

Note that

$$w_k^*(D-1) \le w_k(D) \le w_k^*(D). \tag{3}$$

The upper bound describes the maximum width of circuits in which fanout $k+1$ is allowed as an exception for the nodes v_d of the principal chain. The lower bound describes the width of circuits in which the fanout of nodes v_d is bounded by two. In particular, $w_2(D) = w_2^*(D-1)$.

Claim. *Let $d, h > 0$ and $l = \min\{d, k-1\}$. Then*

$$w_k(d,h) = \sum_{i=1}^{l-1} w_k(d-i, h-1) + 2w_k(d-l, h-1). \tag{4}$$

$\triangleright$ We continue using Fig. 1 as an illustration. Let it depict a pair of consistent trees I and Q with root nodes u_d and v_d, respectively. The immediate descendants of node v_d determine a partition of the index interval $[t,\ t+w]$ into subintervals defined by the indices of the nodes of the chain connecting x_{t+1} and u_d, and also a partition of both trees into pairs of consistent subtrees $(I_j,\ Q_j)$. Consequently,

$$w_k(d,h) = w_k(d_1, h-1) + \ldots + w_k(d_r, h-1), \quad d > d_1 > d_2 > \ldots > d_{r-1} \ge d_r. \tag{5}$$

Obviously, for any d, h,

$$w_k(d, h) \le w_k(d + 1, h). \tag{6}$$

Let us check that for $d \ge 1$ and any h, we also have

$$w_k(d, h) \le 2w_k(d - 1, h). \tag{7}$$

The argument is illustrated in Fig. 2. Consider a pair of consistent trees of width $w = w_k(d, h)$, consisting of a binary tree I of depth $\le d$ and a k-ary tree Q of depth $\le h$. Let y denote the closest ancestor node of root u of tree I lying on the path from the first input x_1. Let subtree I_1, rooted at y, have width τ. Let I_2 denote the subtree of tree I whose leaves are the remaining $w - \tau$ inputs. By the consistency property, the second tree Q contains a node z that is an ancestor of exactly $w - \tau$ higher outputs. Let Q_2 denote the subtree rooted at node z, and Q_1 denote the tree obtained from Q by removing subtree Q_2.

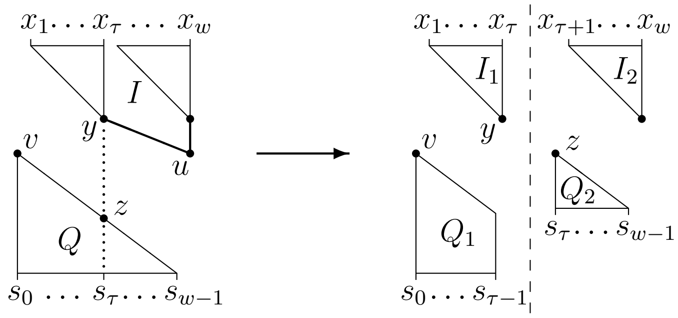

Figure 2: Transformation of a pair of consistent trees

By construction, the pairs of trees (I_1, Q_1) and (I_2, Q_2) are consistent. The total width of the pairs is w. Moreover, the depth of trees I_1, I_2 does not exceed $d - 1$, and the depth of trees Q_1, Q_2 does not exceed h. Hence, (7) is proved.

Now (4) immediately follows from (5) by applying rules (6), (7) and taking into account $r \le k$ and $d_r \ge 0$. □

We proceed directly to the proof of the theorem. Let us estimate $w_k^*(D)$. In view of (4), for $D \ge k$ we have

$$w_k^*(D) = w_k(D, 0) + \sum_{i=2}^{k-1} w_k^*(D - i) + 2w_k^*(D - k) + w_k(0, D) + \sum_{i=2}^{k-1} w_k(0, D - i).$$

Since $w(0, h) = w(d, 0) = 1$ for any $d, h \ge 0$, we obtain the recurrence relation

$$w_k^*(D) = \sum_{i=2}^{k-1} w_k^*(D - i) + 2w_k^*(D - k) + k.$$

This relation, given the initial values $w_k^*(0), \ldots, w_k^*(k-1)$, is resolved in the standard way as $w_k^*(D) \sim c \cdot \alpha_k^D$, where c is some constant, since α_k has the largest absolute value among the roots of polynomial $P_k(x)$: indeed, the modulus of an arbitrary root x satisfies the inequality

$$|x|^k \leq 2 + |x| + |x|^2 + \ldots + |x|^{k-2},$$

whence $|x| \leq \alpha_k$. The assertion of the theorem now follows from (3).

References

[1] Fich F. E. New bounds for parallel prefix circuits. Proc. 15th STOC (Boston, 1983). NY: ACM, 1983, 100–109.

[2] Lin Y.-C., Hung L.-L. Straightforward construction of depth-size optimal, parallel prefix circuits with fan-out 2. ACM Trans. on Design Autom. of Electr. Systems. 2009. **14**(1), Art. 15.

[3] Sheeran M., Parberry I. A new approach to the design of optimal parallel prefix circuits. TR–2006–1, Chalmers Univ. of Technology, 2006.

[4] Snir M. Depth-size trade-offs for parallel prefix computation. J. Algorithms. 1986. **4**, 185–201.

[5] Zhu H., Cheng C.-K., Graham R. On the construction of zero-deficiency parallel prefix circuits with minimum depth. ACM Trans. on Design Autom. of Electr. Syst. 2006. **11**(2), 387–409.