

ISSN 2949-5598

ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ

Том 31 № 4 2024

Новосибирск
Издательство Института математики

О СЛОЖНОСТИ ПАРАЛЛЕЛЬНЫХ ПРЕФИКСНЫХ СХЕМ С ОГРАНИЧЕННЫМ ВЕТВЛЕНИЕМ

И. С. Сергеев

Научно-исследовательский институт «Квант»,
4-й Лихачёвский пер., 15, 125438 Москва, Россия
E-mail: isserg@gmail.com

Аннотация. Доказано, что сложность универсальной префиксной схемы глубины n на 2^n входах с ограничением 2 на ветвление выходов элементов не меньше $0,75(n-1)2^n$. Также приводятся несколько простых конструкций и верхних оценок сложности префиксных схем с ветвлением 2 и глубиной $n+k$. Ил. 4, библиогр. 14.

Ключевые слова: параллельная префиксная схема, схема с ограниченным ветвлением, сложность схемы, глубина схемы.

Введение

Пусть $(S, \circ)$ — полугруппа. Множество функций

$$x_1 \circ x_2 \circ \dots \circ x_i, \quad 1 \leq i \leq m, \quad (1)$$

называется *системой префиксных сумм* переменных $x_1, \dots, x_m$, принимающих значения в S . Схемы из функциональных элементов над базисом $\{x \circ y, x\}$, реализующие систему (1), называют *префиксными схемами*. Напомним, что *сложностью* и *глубиной* схемы называются соответственно общее число элементов схемы и максимальное число элементов в ориентированном пути, соединяющем вход и выход схемы. Подробнее о понятиях схемы, сложности и глубины см., например, в [1, 2]. (Примеры префиксных схем показаны на рис. 1.) Под *ветвлением* вершины в графе (или элемента в схеме) далее понимаем число рёбер, исходящих из неё.

Префиксные схемы применяются во многих теоретических и прикладных задачах синтеза, например в конструкциях сумматоров двоичных чисел, а также в задачах сортировки, решения линейных рекуррентных последовательностей. В ряде задач возникает потребность в построении *параллельных* префиксных схем, т. е. имеющих глубину $O(\log m)$ относительно числа входов m .

Универсальные префиксные схемы правильно вычисляют суммы независимо от выбора полугруппы S ¹⁾. Легко проверить, что избыточная²⁾ универсальная схема использует только операции вида $s_1 \circ s_2$, где $s_1 = x_i \circ x_{i+1} \circ \dots \circ x_j$ и $s_2 = x_{j+1} \circ x_{j+2} \circ \dots \circ x_k$.

Пусть $L(P)$ означает сложность схемы P , а $L_f(k, m)$ — сложность минимальной универсальной префиксной схемы m переменных и глубины $\lceil \log_2 m \rceil + k$ с ограничением f на ветвление входов схемы и внутренних элементов и ограничением $f - 1$ на ветвление выходов схемы³⁾. Положим $L_f(m) = L_f(0, m)$ и $L() = L_\infty()$. Далее, как правило, ограничиваемся рассмотрением наиболее интересного и важного случая $m = 2^n$.

Разнообразные конструкции параллельных префиксных схем предлагались с 1960-х гг., но только около 1980 г. Ладнер и Фишер [3] показали, что можно строить схемы минимальной глубины и линейной сложности. В частности, ими было доказано, что $L(2^n) < 4 \cdot 2^n$. Вскоре Фич [4] улучшила верхнюю оценку и получила нетривиальную нижнюю: $L(2^n) \geq (10/3 - o(1)) \cdot 2^n$. Автор [5, 6] установил точное значение сложности, имеющее вид $L(2^n) = (3,5 - o(1)) \cdot 2^n$. Лучшие верхние оценки для схем почти минимальной глубины также получены в [5, 6]. Они имеют вид $L(k, 2^n) \leq (2 + 2^{-k})2^n - O(2^{(n-k)/2})$ при $1 \leq k \ll n$. При $k \geq (\log_\varphi 2 - 1)n - O(1) \approx 0,44n - O(1)$, где $\varphi = \frac{1+\sqrt{5}}{2}$, выполнено $L(k, 2^n) = 2^{n+1} - n - k - 2$ (см. [7]). Более подробное изложение этой линии результатов см. в [6].

С точки зрения электроники существенное влияние на характеристики схемы, помимо глубины и сложности, оказывает также степень ветвления элементов: чем степень меньше, тем лучше. Поэтому значительный объём прикладных и теоретических исследований посвящён параллельным префиксным схемам с ограниченным ветвлением. Отметим, что только ограничение 2 на ветвление принуждает к использованию тождественных элементов. В других случаях без них можно обойтись.

Изящную конструкцию схем минимальной глубины и ветвления 2 предложили в начале 1970-х гг. Коге и Стоун [8]. Их пример доказывает оценку

$$L_2(2^n) \leq (n - 0,5)2^n. \quad (2)$$

Принципиальный результат получила Фич [4], показав, что линейная сложность для схем минимальной глубины с ветвлением 2 недостижима:

$$L_2(2^n) \geq (n + 1)2^{n-1} - O(n^2). \quad (3)$$

¹⁾ Достаточно потребовать, чтобы результат был правильным, скажем, для группы симметрий окружности (см. [6]).

²⁾ Т. е. не содержащая элементов, не связанных с выходами.

³⁾ Более жёсткое условие на ветвление выходов накладывается, чтобы обеспечить возможность использования префиксных схем в качестве подсхем.

С другой стороны, при любом $f \geq 3$ выполнено $L_f(2^n) = O(2^n)$. Фич указала способ построения таких схем⁴⁾.

Ослабляя ограничение на глубину, можно добиться линейной сложности и для схем с ветвлением 2. Подобную схему построили Брент и Кунг [9]. Их конструкция влечёт оценку $L_2(n-1, 2^n) \leq 2,5 \cdot 2^n - n - 3$. Используя рекурсивный метод [3], но выбирая схему Коге — Стоуна в качестве базовой, Хан и Карлсон [10] построили семейство схем, приводящее при $1 \leq k \leq n-1$ к оценке сложности⁵⁾

$$L_2(k, 2^n) \leq (n-k-3)2^{n-k} + 5 \cdot 2^{n-1} - k. \quad (4)$$

Автор [6] предложил модификацию базовой схемы, уточнив оценку сложности при $1 \leq k \leq n-2$ до

$$L_2(k, 2^n) \leq (n-k-3,25)2^{n-k} + 5 \cdot 2^{n-1} - k. \quad (5)$$

Схемы из двух семейств отличаются только числом тождественных элементов. Нижняя оценка из [4] имеет вид

$$L_2(k, 2^n) \geq (n-2)2^{n-k-1} + 2^n - O(n^2 2^{-k})$$

и содержательна только при $k \ll \log_2 n$, поскольку при больших k приоритет имеет оценка $L(k, 2^n) \geq 2^{n+1} - n - k - 2$, вытекающая из стандартного соотношения

$$L + D \geq 2m - 2, \quad (6)$$

справедливого для произвольной префиксной схемы m переменных глубины D ⁶⁾ с числом двухвходовых элементов L [4, 11]. Схемы, на которых достигается равенство в (6), называют *оптимальными*.

Схемы линейной сложности с ветвлением $f \geq 3$, построенные Фич [4], не вполне практичны, поскольку мультипликативные константы в оценках велики: например, $L_3(2^n) \leq 54,25 \cdot 2^n$. В современной работе [12] авторы путём модификации семейства схем Ладнера — Фишера получили более аккуратные оценки, в частности $L_3(2^n) \leq (7,75 + o(1)) \cdot 2^n$.

Имеется обширная литература по параллельным префиксным схемам с различными ограничениями на ветвление элементов. Например, в работах [13, 14] построены оптимальные параллельные префиксные схемы с ветвлением 2 и максимально возможным числом входов (там же приводится подробная библиография по этому направлению). Минимально возможная глубина оптимальной схемы с ветвлением 2 на m входах

⁴⁾ Она также получила верхние оценки сложности схем с ветвлением 2, но они, во-первых, превышают сложность схем Коге — Стоуна, а во-вторых, демонстрируются на схемах, некоторые входы которых имеют ветвление 3.

⁵⁾ В качестве частного случая при $k = n-1$ получается схема Брента — Кунга.

⁶⁾ Более точно, оценка доказывается для D , равного глубине вычисления старшей префиксной суммы — суммы всех переменных — в схеме.

оказывается равной $\lfloor \log_2 m \rfloor + \lfloor \log_2(2m/3) \rfloor$. Однако общая сложность схем с учётом тождественных элементов авторами не оценивалась.

Далее докажем нижнюю оценку $L_2(2^n) > \frac{3}{4}(n-1)2^n$, сокращая разрыв между известными верхней (2) и нижней (3) оценками сложности параллельных префиксных схем с ветвлением 2. Новая оценка опирается на свойство минимальных схем вычислять только «интервальные» суммы вида $x_l \circ x_{l+1} \circ \dots \circ x_r$, поэтому справедлива не только для универсальных схем, но и для схем над конкретными некоммутативными полугруппами, включая идемпотентные. В частности, она применима в случае полугруппы, используемой при построении (префиксных) сумматоров⁷⁾. Отметим, что предшествующая оценка Фич выполняется для схем над любой конкретной полугруппой, за исключением вырожденных случаев⁸⁾.

В заключение сделаем несколько простых замечаний относительно верхних оценок сложности. Сначала покажем, что оценки (4), (5) могут быть уточнены на величину $O(2^k)$. Далее, докажем, что асимптотически минимально возможная для параллельной схемы оценка $L_2(k, m) = (2 + o(1))m$ достигается при $k = O(\log \log m)$, а при $k = (1 + o(1)) \log_2 m$ демонстрируется оптимальными префиксными схемами.

1. Нижняя оценка

Лемма 1. В корневом бинарном дереве с m листьями минимально возможная сумма длин всех путей от корня к листьям равна $(l+1)m - 2^l$, где $l = \lfloor \log_2 m \rfloor$.

ДОКАЗАТЕЛЬСТВО. Это решение известной оптимизационной задачи, связанной с неравенством Крафта — Макмиллана и алфавитным кодированием. Задача состоит в минимизации суммы (глубин листьев) $\sum_{i=1}^m d_i$ при условии (существования дерева с заданным набором глубин)

$$\sum_{i=1}^m 2^{-d_i} \leq 1. \quad (7)$$

Несложно видеть, что решение доставляет набор чисел d_i , которые принимают одно или два соседних целочисленных значения. Действительно, если, скажем, $d_1 \leq d_2 + 2$, то пару чисел d_1, d_2 можно заменить парой $d_1 + 1, d_2 - 1$. При этом целевая сумма не изменится, а неравенство (7) останется в силе.

В оптимальном дереве $2^l - m$ вершин расположены на глубине $l - 1$ и $2m - 2^l$ — на глубине l . Лемма 1 доказана.

⁷⁾ Имеется в виду множество $\{0, 1\}^2$ с операцией $(f, g) \circ (f', g') = (f \vee gf', gg')$.

⁸⁾ Скажем, когда сумма $a \circ b$ не зависит от одного из слагаемых.

Далее для записи интервальных префиксных сумм будем использовать сокращённое обозначение $x_{[l;r]} = x_l \circ x_{l+1} \circ \dots \circ x_r$.

Напомним, что в избыточной универсальной префиксной схеме любой элемент вычисляет сумму вида $x_{[l;r]}$. Число r назовём *индексом* элемента.

Теорема 1. Справедливо $L_2(2^n) > \frac{3}{4}(n-1)2^n$.

Доказательство. При $n \geq 2$ рассмотрим избыточную префиксную схему Π глубины n на 2^n входах. Схему можно условно разбить на две половины: в левой — элементы с индексами из $[1, 2^{n-1}]$, в правой — с индексами из $[2^{n-1} + 1, 2^n]$. Обозначим через $\lambda(\Pi)$ и $\rho(\Pi)$ число элементов в левой и правой половинах схемы соответственно.

Следующее наблюдение обычно для анализа параллельных префиксных схем минимальной глубины безотносительно ограничений на ветвление элементов.

Утверждение 1. $\rho(\Pi) \geq L_2(2^{n-1}) + 2^{n-1}$.

Доказательство. Заметим, во-первых, что в правой половине схемы на глубине n расположены 2^{n-1} элементов, вычисляющих финальные суммы $x_{[1;i]}$, $2^{n-1} < i \leq 2^n$, как $s_i^0 \circ s_i^1$, где подсумма s_i^0 вычисляется в левой половине схемы, а s_i^1 — в правой на глубине не выше $n-1$.

Оборвём рёбра, соединяющие элементы из левой половины с элементами из правой. Тогда элементы, прежде вычислявшие суммы s_i^1 , будут вычислять суммы $x_{[2^{n-1}+1;i]}$, т. е. все возможные префиксные суммы переменных $x_{2^{n-1}+1}, x_{2^{n-1}+2}, \dots, x_{2^n}$ с глубиной $n-1$. Степени ветвления элементов в правой половине схемы Π при этом не изменяются. Утверждение 1 доказано.

Теперь оценим снизу $\lambda(\Pi)$. Будем анализировать пути, соединяющие входы x_i , $1 \leq i \leq 2^{n-1}$, с выходами q_j , вычисляющими суммы $x_{[1;j]}$. Для любой пары (i, j) при $i \leq j$ в универсальной схеме Π имеется ровно один путь от x_i до q_j ⁹⁾. Если путь из x_i в q_j , где $i \leq j$, не содержит элемента глубины d , то будем говорить, что путь *пропускает уровень d* .

I. Сделаем несколько простых замечаний. Пусть $E_{[r_1, r_2]}^d$ обозначает множество расположенных на глубине d элементов схемы Π с индексами в интервале $[r_1, r_2]$.

⁹⁾ В общем случае путей может быть несколько. Доказательство проходит и в некоторых таких ситуациях, например, в случае идемпотентной некоммукативной операции. Тогда следует выбрать по одному пути для каждого выхода так, чтобы множество выбранных путей образовывало дерево (с корневой вершиной x_1), а остальные пути не принимать в расчёт.

1) Отметим, что любой путь из вершины x_i к выходам либо проходит через элемент из $E_{[i, i+2^d-1]}^d$, либо пропускает уровень d .

2) Через любой элемент (в силу ограничения на ветвление) проходит не более 2^n различных путей, соединяющих входы и выходы схемы: элемент на глубине d связан не более чем с 2^d входами и не более чем с 2^{n-d} выходами.

3) Согласно лемме 1 пути от входа x_i , где $i \leq 2^{n-1}$, к выходам (таких путей $2^n - i + 1$) пропускают суммарно не более $i - 1$ уровней (напомним, что всего в схеме n уровней).

II. Обозначим через $\lambda_d = |E_{[1, 2^{n-1}]}^d|$ число элементов в левой половине схемы П на глубине d . Согласно замечаниям из предыдущего пункта через эти элементы проходит не более $\lambda_d 2^n$ путей между входами и выходами схемы. Пути из вершин $x_1, \dots, x_{2^{n-1}-2^d+1}$ могут проходить только через эти элементы, либо пропускать уровень d . Общее число путей из указанных вершин к выходам равно

$$\pi_d = 2^n + (2^n - 1) + \dots + (2^{n-1} + 2^d) = 3(2^{2n-3} + 2^{n-2}) - (2^n - 1)2^{d-1} - 2^{2d-1}.$$

Таким образом, не менее $\sigma_d = \max\{0, \pi_d - \lambda_d 2^n\}$ из этих путей пропускают уровень d .

Оценивая при помощи леммы 1 общее число пропусков уровней путями от входов $x_1, \dots, x_{2^{n-1}-1}$ к выходам, получаем

$$\sum_{d=1}^{n-1} \sigma_d \leq \frac{1}{2}(2^{n-1} - 1)(2^{n-1} - 2) = 2^{2n-3} - 3 \cdot 2^{n-2} + 1.$$

Ввиду оценки $\lambda_d \geq 2^{-n}(\pi_d - \sigma_d)$ получаем

$$\begin{aligned} \lambda(\Pi) &\geq \sum_{d=1}^{n-1} \lambda_d \geq (3n - 4)2^{n-3} + \frac{3n}{4} - 2^{-n} - \\ &\quad - (1 - 2^{-n}) \sum_{d=1}^{n-1} 2^{d-1} - 2^{-n} \sum_{d=1}^{n-1} 2^{2d-1} = \\ &= 3n2^{n-3} - \frac{7}{6} \cdot 2^n + O(n). \quad (8) \end{aligned}$$

При помощи утверждения 1 отсюда уже извлекается оценка сложности в виде $L_2(2^n) \geq \frac{3}{4} \cdot n2^n - O(2^n)$. В следующем пункте уточним линейное слагаемое (порядка 2^n) в (8), чтобы получить чуть более точную оценку, указанную в формулировке теоремы.

III. При $n \geq 4$ оценим более аккуратно число элементов на уровнях $n - 3$, $n - 2$, $n - 1$. Запишем

$$\lambda_{n-2} = \lambda_{n-2,0} + \lambda_{n-2,1}, \quad \lambda_{n-3} = \lambda_{n-3,0} + \dots + \lambda_{n-3,3},$$

где $\lambda_{n-k,i} = |E_{[i2^{n-k}+1,(i+1)2^{n-k}]}^{n-k}|$. Напомним, что элемент на глубине $n-k$ связан путями не более чем с 2^k выходами схемы.

Все пути из x_1 к выходам проходят через элементы из $E_{[1,2^d]}^d$, этих путей 2^n , они не пропускают уровни, поэтому $\lambda_{n-1} = 2^{n-1}$, $\lambda_{n-2,0} = 2^{n-2}$, $\lambda_{n-3,0} = 2^{n-3}$.

Пути из $x_{2^{n-2}+1}$ проходят через элементы из $E_{[2^{n-2}+1,2^{n-2}+2^d]}^d$ либо пропускают соответствующие уровни, всего путей $3 \cdot 2^{n-2}$, а число пропусков не превосходит 2^{n-2} , поэтому

$$\max\{3 \cdot 2^{n-2} - 4\lambda_{n-2,1}, 0\} + \max\{3 \cdot 2^{n-2} - 8\lambda_{n-3,2}, 0\} \leq 2^{n-2},$$

откуда $\lambda_{n-2,1} + \lambda_{n-3,2} \geq 7 \cdot 2^{n-5}$.

Наконец, пути из вершины $x_{i2^{n-3}+1}$ проходят через $E_{[i2^{n-3}+1,(i+1)2^{n-3}]}^{n-3}$ или пропускают уровень $n-3$, таких путей $(8-i)2^{n-3}$, а пропускают они суммарно не более $i2^{n-3}$ уровней, поэтому получаем $\lambda_{n-3,1} \geq 3 \cdot 2^{n-5}$ и $\lambda_{n-3,3} \geq 2^{n-5}$.

Окончательно находим: $\lambda_{n-1} + \lambda_{n-2} + \lambda_{n-3} \geq 39 \cdot 2^{n-5}$. Теперь оценка (8) усиливается до следующей:

$$\begin{aligned} \lambda(\Pi) &\geq \sum_{d=1}^{n-1} \lambda_d \geq 39 \cdot 2^{n-5} + 2^{-n} \sum_{d=1}^{n-4} \pi_d - 2^{-n} \sum_{d=1}^{n-1} \sigma_d \geq \\ &\geq 39 \cdot 2^{n-5} + (3n-13)2^{n-3} + \frac{3n-9}{4} - 2^{-n} - \\ &\quad - (1-2^{-n}) \sum_{d=1}^{n-4} 2^{d-1} - 2^{-n} \sum_{d=1}^{n-4} 2^{2d-1} = \\ &= 3n2^{n-3} - \frac{181}{192} \cdot 2^{n-1} + \frac{3n}{4} - \frac{19}{16} - \frac{2^{2-n}}{3} > 3n2^{n-3} - 2^{n-1}, \end{aligned}$$

так как внутренние суммы в третьей строке равны $2^{n-4} - 1$ и $\frac{2}{3}(4^{n-4} - 1)$.

IV. Применяя утверждение 1, приходим при $n \geq 4$ к рекуррентному соотношению

$$L_2(2^n) > L_2(2^{n-1}) + \frac{3}{4}n2^{n-1}.$$

В качестве базы индукции возьмём достаточно несложно получаемую оценку¹⁰⁾ $L_2(8) \geq 13$. Теперь соотношение разрешается так, как заявлено в формулировке теоремы. Результат верен и в случае $0 \leq n \leq 3$. Теорема 1 доказана.

¹⁰⁾ Например, она следует из легко проверяемого соотношения $L_2(4) = 6$ и утверждения 1, применяемого с $n = 3$. На самом деле, рассуждение п. III вместе с утверждением 1 приводят к оценке $L_2(8) \geq 17$.

2. Некоторые верхние оценки

2.1. Напомним способ, которым строятся схемы почти минимальной глубины с ветвлением 2, дающие оценки сложности (4) и (5). Указанные конструкции основаны на базовых схемах Коге — Стоуна [8] минимальной глубины и их модификации из работы [6]. Схемы этих семейств обозначаем через K_n и S_n соответственно, где 2^n — число входов. На рис. 1 показаны схемы с 8 входами.

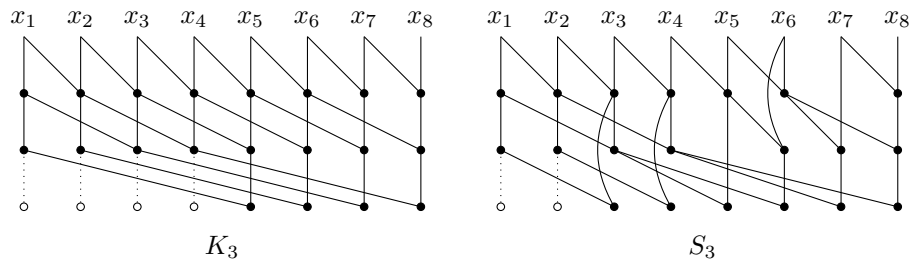


Рис. 1. Схема Коге — Стоуна (слева) и схема из [6] (справа)

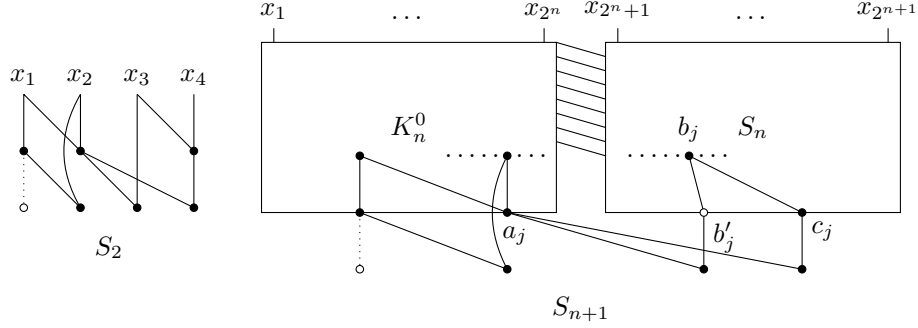
При изображении префиксных схем следуем традиции располагать входы сверху слева направо, элементы с одинаковым индексом — друг под другом, элементы одного уровня — на одной горизонтали. Элементы обозначаются чёрными кружками.

В рекурсивных построениях применяются варианты базовых схем, достроенные до схем с нулевым ветвлением выходов. На рис. 1 дополнительные элементы изображаются белыми кружками, а соединительные линии — пунктиром. Таким образом достроенные варианты схем будем обозначать через K_n^0 и S_n^0 .

Схемы Коге — Стоуна K_n имеют простую регулярную структуру. Поясним строение схемы S_n на рис. 2. Схема S_{n+1} составлена из подсхем K_n^0 и S_n , которые стыкуются на первых $n - 1$ уровнях по правилу схем Коге — Стоуна. В результате стыковки элементы b_j на уровне $n - 1$ в левой половине подсхемы S_n вычисляют суммы $x_{[2^{n-1}+j+1; 2^n+j]}$, $j = 1, \dots, 2^{n-1}$. Элемент b_j соединён с одним (c_j) или двумя¹¹⁾ (b'_j и c_j) выходами подсхемы S_n . Выход a_j подсхемы K_n^0 , в котором вычисляется сумма $x_{[1; 2^{n-1}+j]}$, и элементы b_j (либо b'_j) и c_j участвуют в вычислении пары финальных префиксных сумм. Сумма $x_{[1; 2^{n-1}+j]}$ вычисляется повторно на следующем уровне.

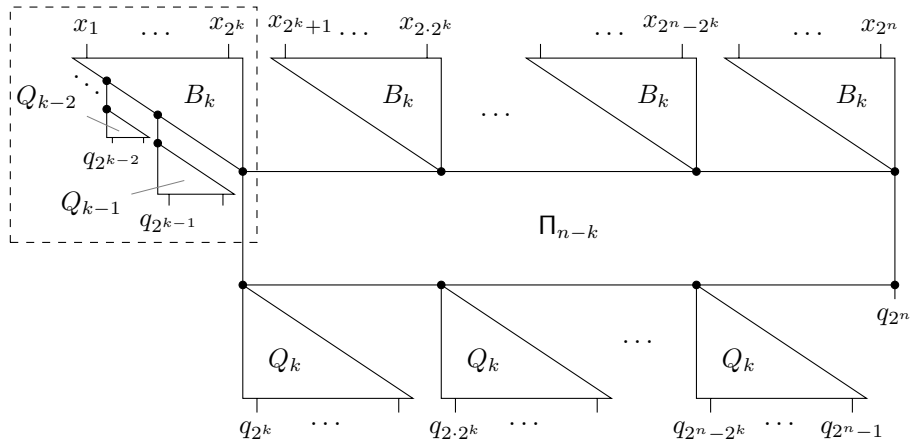
Сложность схем K_n и S_n совпадает, $L(K_n) = L(S_n) = (n - 0,5)2^n$, при этом дополненные схемы S_n^0 имеют меньшую сложность, чем схемы K_n^0 ,

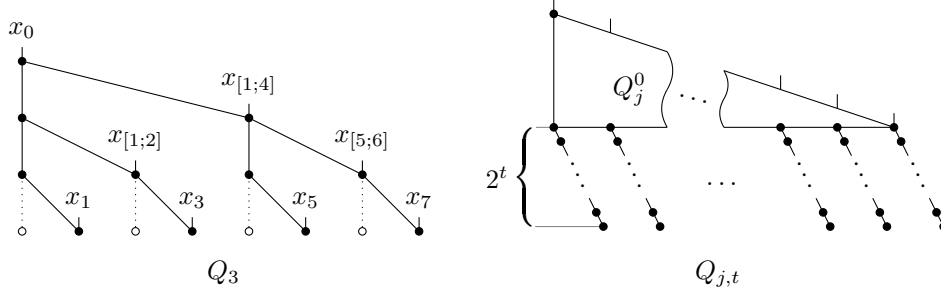
¹¹⁾ С одним при $j \leq 2^{n-2}$ и с двумя при $j > 2^{n-2}$.

Рис. 2. Рекурсивное определение схем S_n

а именно, $L(K_n^0) = n2^n$ и $L(S_n^0) = (n - 0,25)2^n$ (разница в числе тождественных элементов).

В методе из [10] (который по сути является переложением метода из [3] на схемы с ограничением ветвления) строятся схемы $\Pi_{k,n}$ глубины $n+k$ на 2^n входах, которые имеют вид, изображённый на рис. 3. Входы разбиты на группы по 2^k штук. Суммы в группах вычисляются полными бинарными деревьями B_k . Их выходы присоединены к префиксной схеме Π_{n-k} минимальной глубины. На выходах этой подсхемы реализуются все префиксные суммы с числом слагаемых, кратным 2^k . Полный комплект сумм исходных переменных вычисляется присоединёнными «корневыми» вершинами к выходам внутренней подсхемы «перевёрнутыми» бинарными деревьями Q_k , а также деревьями $Q_1, Q_2, \dots, Q_{k-1}$,

Рис. 3. Схема $\Pi_{k,n}$

Рис. 4. Конструкции (под)схем Q_j и $Q_{j,t}$

присоединёнными к внутренним элементам главной цепочки¹²⁾ первого дерева B_k . Другими входами подсхем Q_j являются переменные и промежуточные суммы, вычисляемые на первых $k-1$ уровнях. На рис. 4 слева изображён пример схемы Q_3 (с 8 выходами). На выходах q_j схемы $\Pi_{k,n}$ вычисляются суммы $x_{[1;j]}$.

На схемах $\Pi_{k,n}$ достигаются оценки сложности (4) или (5), в зависимости от того, какая базовая схема используется. Отметим, что они удовлетворяют ещё одному существенному ограничению: старшая префиксная сумма вычисляется на глубине n .

Заметим, что обведённый пунктиром (левый верхний) фрагмент схемы $\Pi_{k,n}$ имеет запас по глубине и поэтому реализован неоптимально. Чтобы получить небольшой выигрыш в сложности, вместо схем Q_j для вычисления финальных сумм в рассматриваемом фрагменте можно использовать схемы типа $Q_{j,t}$. Схема $Q_{j,t}$ функционально эквивалентна схеме Q_{j+t} . Она позволяет получить 2^{j+t} выходных сумм, но имеет глубину $j + 2^t - 1$ (см. рис. 4 (справа)). Внутренняя подсхема Q_j^0 (т. е. схема Q_j , дополненная до схемы с нулевым ветвлением выходов) вычисляет суммы с числом слагаемых, кратным 2^t . Остальные суммы вычисляются последовательно прибавлением одной переменной к уже вычисленным. Схема $Q_{j,1}$ совпадает с Q_{j+1} .

Без учёта корневой вершины схемы Q_{j+t} и $Q_{j,t}$ содержат по $2^{j+t} - 1$ бинарных элементов, но $2^{j+t-1} - 1$ и $2^j - 1$ тождественных элементов соответственно.

Таким образом, замена подсхем Q_j в выделенной части схемы $\Pi_{k,n}$ подсхемами $Q_{j-t,t}$ с максимально возможным t позволяет при $k \geq 3$ уточнить оценки (4), (5) на величину $O(2^k)$ за счёт сокращения некоторых тождественных элементов.

¹²⁾ Главная цепочка соединяет вход x_1 с выходом, в котором вычисляется старшая префиксная сумма. В данном случае элементы цепочки вычисляют суммы $x_{[1;2]}, x_{[1;4]}, \dots, x_{[1;2^{k-1}]}$.

При малых k экономия незначительна, однако, например, сложность схемы Брента — Кунга ($k = n - 1$) только при замене $Q_3, Q_4, \dots, Q_{k-1}$ на $Q_{1,2}, Q_{2,2}, \dots, Q_{k-3,2}$ уменьшается на $2^{n-3} - 1$ до $2^{\frac{3}{8}} \cdot 2^n - n - 2$ и далее ещё примерно на $0,01 \cdot 2^n$ при использовании подсхем $Q_{j,t}$ с $t \geq 3$.

Если дублировать элементы в главной цепочке (ограничение на ветвление это позволяет), то вообще все внешние подсхемы $Q_3, Q_4, \dots, Q_k$ можно заменить подсхемами $Q_{j,t}$ с $t \geq 2$, что приводит к схеме сложности менее $2,25 \cdot 2^n$, но при этом число бинарных элементов увеличивается на $n - 2$.

Отказ от (несколько искусственного) ограничения на глубину вычисления старшей префиксной суммы позволил бы перейти к ещё более экономным конструкциям. Далее сформулируем несколько общих оценок.

2.2. Как следует из (6), число L бинарных элементов в префиксной схеме с m входами, имеющей глубину D (или глубину D вычисления старшей суммы $x_{[1;m]}$), никогда не меньше чем $2m - D - 2$. Оценка достигается, если D не слишком мало, в частности в случае ограничения 2 на ветвление элементов, при

$$D \geq \lfloor \log_2 m \rfloor + \lfloor \log_2(2m/3) \rfloor = 2 \log_2 m - O(1).$$

Соответствующие конструкции схем приведены в [13, 14]. Фактически из процедуры построения следует доказательство минимальности глубины для класса оптимальных схем. Схема [9] оптимальна в слабом смысле, т. е. при дополнительном ограничении на глубину вычисления старшей суммы, и также имеет минимально возможную для оптимальной схемы глубину.

Приведём несколько простых результатов для общей сложности. Множество элементов любой префиксной схемы можно разбить на оригиналы и дубликаты. Суммы, вычисляемые *оригиналами*, попарно различны. *Дубликат* вычисляет сумму, которая либо тождественна одной из переменных, либо встречается среди сумм, вычисляемых оригиналами. Легко проверить, что число дубликатов в любой схеме с ветвлением 2 по порядку не меньше m/D . Справедлива

Лемма 2. Число дубликатов в префиксной схеме глубины D с m входами и ветвлением 2 не меньше $\frac{m-D-1}{2(D-1)}$.

Доказательство. Обозначим через U множество всех элементов избыточной схемы, в которых вычисляются суммы $x_{[1;i]}$. Через U_2 обозначим множество всех бинарных элементов в U . Заметим, что бинарные элементы вычисляют все $m - 1$ префиксных сумм $x_{[1;i]}$, кроме x_1 .

Множество U_2 естественным образом распадается на непересекающиеся ориентированные цепочки с выколотыми стартовыми вершинами: такие вершины принадлежат U , но не обязательно U_2 . Опишем подходящее

неприводимое разбиение — такое, в котором никакие две цепочки нельзя объединить в одну.

Будем строить цепочки в порядке возрастания глубины стартовых вершин до тех пор, пока множество U_2 не будет исчерпано. Формируя очередную цепочку, продолжаем её до тех пор, пока к ней нельзя будет добавить новую вершину из U_2 . Главная цепочка (она единственная) начинается от входа x_1 . Стартовой вершиной любой очередной цепочки может быть либо тождественный элемент, либо бинарный элемент ранее построенной цепочки. В первом случае поставим цепочке в соответствие стартовый тождественный элемент — можем считать его дубликатом. Во втором случае ветвление стартового элемента в схеме равно 2, поэтому он не может быть выходом, и, как следствие, для него найдётся дубликат. Среди всех дубликатов (вычисляющих одну и ту же сумму) выберем и поставим в соответствие построенной цепочке тот, который 1) не был ранее сопоставлен цепочке с бинарным стартовым элементом и 2) является либо бинарным элементом, либо выходом. Условиям 1 и 2 всегда можно удовлетворить, поскольку среди элементов, вычисляющих фиксированную сумму $x_{[1;j]}$, бинарных элементов с ветвлением 2 (т. е. тех, которые могут быть стартовыми вершинами) меньше, чем тех, которые удовлетворяют условию 2.

По построению каждый элемент-дубликат соответствует не более чем двум цепочкам разбиения. В главной цепочке вычисляется не более D выходных сумм, ей не ставятся в соответствие дубликаты. Каждая очередная цепочка добавляет не более чем $D - 1$ новых выходных сумм. Лемма 2 доказана.

Дубликаты — это цена, которую приходится платить за ограничение ветвления. Чаще всего дубликатами являются тождественные элементы¹³⁾.

Теорема 2. Соотношение $L_2(k, m) = (2 + o(1))m$ достигается

- 1) при $k = O(\log \log m)$;
- 2) на оптимальной префиксной схеме при $k = (1 + o(1)) \log_2 m$.

ДОКАЗАТЕЛЬСТВО. 1) Пусть $m = t2^n$. Разобьём входные переменные на группы по t штук и вычислим групповые суммы $y_i = x_{[(i-1)t+1; it]}$, $1 \leq i \leq 2^n$, с глубиной $h = \lceil \log_2 t \rceil$ и общей сложностью $(t - 1)2^n$.

На входах y_i построим префиксную схему глубины $n + \log_2 n$ и линейной сложности $O(2^n)$. Такая схема существует, например, согласно (5). На выходах схемы вычислены все суммы $x_{[1; it]}$, $1 \leq i \leq 2^n$.

¹³⁾ При условии, что оригиналами могут быть только бинарные элементы.

Оставшиеся суммы вычислим для каждой группы переменных при $i = 0, \dots, 2^n - 1$ последовательно как

$$x_{[1;it+1]} = x_{[1;it]} \circ x_{it+1}, \quad \dots, \quad x_{[1;it+t-1]} = x_{[1;it+t-2]} \circ x_{it+t-1},$$

используя при необходимости тождественный элемент в начале каждой цепочки¹⁴⁾. Эти вычисления выполняются с глубиной t и сложностью $t2^n$.

Общая сложность схемы равна $2m + O(2^n)$, а глубина — $n + \log_2 n + t + h$. Нужный результат получается, например, при $t \sim \log_2 n$. Конструкция легко адаптируется на случай произвольного m .

2) Этот пункт доказывается аналогично. Напомним, что оптимальность префиксной схемы заключается в выполнении следующих условий [4, 7]:

- все её бинарные элементы либо принадлежат дереву, вычисляющему старшую сумму $x_{[1;m]}$, либо вычисляют выходные суммы $x_{[1;i]}$ — по одному разу каждую;
- старшая сумма вычисляется на глубине, равной глубине схемы.

Положим $m = t(2^n + 1) + n$. На первом шаге вычисляются групповые суммы $y_i = x_{[(i-1)t+1;it]}$, $1 \leq i \leq 2^n$, с глубиной h и общей сложностью $(t-1)2^n$.

К выходам y_i присоединим схему Брента — Кунга¹⁵⁾: она имеет глубину $2n - 1$, сложность около $2,5 \cdot 2^n$, а старшую сумму $y_{[1;2^n]}$ вычисляет с глубиной n .

Как и в конструкции из п. 1, оставшиеся суммы вычислим по группам последовательно. При этом добавим в схему 2^n цепочек: каждая состоит из $t - 1$ бинарных элементов и начинается, при необходимости, с тождественного элемента. Старшие суммы $x_{[1;t2^n+1]}, \dots, x_{[1;t2^n+t+n]}$ вычисляем в цепочке длины $n + t$, присоединённой к старшему выходу подсхемы Брента — Кунга.

Таким образом, общая сложность схемы равна $2m + O(2^n)$, а глубина равна $h + 2n + t < 2\log_2 m + t$ и совпадает с глубиной вычисления старшей суммы. По построению множество бинарных элементов удовлетворяет условию оптимальности. Нужный результат получается при выборе параметра t медленно растущим, скажем, $t \in \omega(1) \cap o(n)$. Теорема 2 доказана.

Заметим, что при выборе $t = \Theta(n)$ в любой из конструкций теоремы 2 одновременно достигается минимальное по порядку число дубликатов $\Theta(m/\log m)$ для параллельной префиксной схемы (согласно лемме 2),

¹⁴⁾ Если цепочка присоединяется к элементу, имеющему ветвление 1.

¹⁵⁾ Можно использовать и схемы из [13, 14].

но ценой дополнительного увеличения глубины. В оптимальной схеме дубликаты составляют в точности множество тождественных элементов.

Финансирование работы

Исследование выполнено за счёт бюджета Научно-исследовательского института «Квант». Дополнительных грантов на проведение или руководство этим исследованием получено не было.

Конфликт интересов

Автор заявляет, что у него нет конфликта интересов.

Литература

1. **Лупанов О. Б.** Асимптотические оценки сложности управляющих систем. М.: Изд-во МГУ, 1984. 138 с.
2. **Jukna S.** Boolean function complexity. Heidelberg: Springer, 2012. 618 p.
3. **Ladner R. E., Fischer M. J.** Parallel prefix computation // J. ACM. 1980. V. 27, No. 4. P. 831–838.
4. **Fich F. E.** New bounds for parallel prefix circuits // Proc. 15th ACM Annu. Symp. Theory of Computing (Boston, USA, April 25–27, 1983). New York: ACM, 1983. P. 100–109.
5. **Сергеев И. С.** О минимальных параллельных префиксных схемах // Вестн. Моск. унив. Сер. 1. Математика. Механика. 2011. № 5. С. 48–51.
6. **Sergeev I. S.** On the complexity of parallel prefix circuits // Electron. Colloq. Comput. Complex.: Tech. Rep. TR13-041. Rehovot: Weizmann Inst. Sci., 2013. 47 p.
7. **Zhu H., Cheng C.-K., Graham R.** On the construction of zero-deficiency parallel prefix circuits with minimum depth // ACM Trans. Des. Autom. Electron. Syst. 2006. V. 11, No. 2. P. 387–409.
8. **Kogge P. M., Stone H. S.** A parallel algorithm for the efficient solution of a general class of recurrence equations // IEEE Trans. Comput. 1973. V. 22, No. 8. P. 786–793.
9. **Brent R. P., Kung H. T.** A regular layout for parallel adders // IEEE Trans. Comput. 1982. V. 31, No. 3. P. 260–264.
10. **Han T., Carlson D. A.** Fast area-efficient VLSI adders // Proc. IEEE 8th Symp. Computer Arithmetic (Como, Italy, May 19–21, 1987). Los Alamitos: IEEE, 1987. P. 49–56.
11. **Snir M.** Depth-size trade-offs for parallel prefix computation // J. Algorithms. 1986. V. 4. P. 185–201.
12. **Bund J., Lenzen C., Medina M.** Optimal metastability-containing sorting via parallel prefix computation // IEEE Trans. Comput. 2020. V. 69, No. 2. P. 198–211.
13. **Lin Y.-C., Hung L.-L.** Straightforward construction of depth-size optimal, parallel prefix circuits with fan-out 2 // ACM Trans. Des. Autom. Electron. Syst. 2009. V. 14, No. 1. Paper ID 15. 13 p.

14. **Sheeran M., Parberry I.** A new approach to the design of optimal parallel prefix circuits. Tech. Rep. No. 2006:1. Göteborg: Chalmers Univ. Tech., 2006. 25 p.

Сергеев Игорь Сергеевич

Статья поступила

22 декабря 2023 г.

После доработки —

14 апреля 2024 г.

Принята к публикации

22 июня 2024 г.

ON THE COMPLEXITY OF FANOUT-BOUNDED PARALLEL PREFIX CIRCUITS

I. S. Sergeev

Scientific and Research Institute “Kvant”,
15 Chetvyortyi Likhachyovskii Lane, 125438 Moscow, Russia
E-mail: isserg@gmail.com

Abstract. We prove that the complexity of a universal depth- n parallel prefix circuit on 2^n inputs with fanout bounded by 2 is at least $0.75(n-1)2^n$. We also propose a number of simple constructions and upper complexity bounds on fanout-2 prefix circuits of depth $n+k$. Illustr. 4, bibliogr. 14.

Keywords: parallel prefix circuit, fanout-bounded circuit, circuit complexity, circuit depth.

References

1. **O. B. Lupanov**, *Asymptotic Bounds for the Complexity of Control Systems* (Izd. Mosk. Gos. Univ., Moscow, 1984) [Russian].
2. **S. Jukna**, *Boolean Function Complexity* (Springer, Heidelberg, 2012).
3. **R. E. Ladner** and **M. J. Fischer**, Parallel prefix computation, *J. ACM* **27** (4), 831–838 (1980).
4. **F. E. Fich**, New bounds for parallel prefix circuits, in *Proc. 15th ACM Annu. Symp. Theory of Computing, Boston, USA, April 25–27, 1983* (ACM, New York, 1983), pp. 100–109.
5. **I. S. Sergeev**, Minimal parallel prefix circuits, *Vestn. Mosk. Univ., Ser. 1. Mat. Mekh.*, No. 5, 48–51 (2011) [Russian] [*Mosc. Univ. Math. Bull.* **66** (5), 215–218 (2011)].
6. **I. S. Sergeev**, On the complexity of parallel prefix circuits, *Electron. Colloq. Comput. Complex* (Tech. Rep. TR13-041) (Weizmann Inst. Sci., Rehovot, 2013).
7. **H. Zhu**, **C.-K. Cheng**, and **R. Graham**, On the construction of zero-deficiency parallel prefix circuits with minimum depth, *ACM Trans. Des. Autom. Electron. Syst.* **11** (2), 387–409 (2006).

English transl.: *Journal of Applied and Industrial Mathematics* **18** (4), 851–860 (2024), DOI: 10.1134/S1990478924040185.

8. **P. M. Kogge** and **H. S. Stone**, A parallel algorithm for the efficient solution of a general class of recurrence equations, *IEEE Trans. Comput.* **22** (8), 786–793 (1973).
9. **R. P. Brent** and **H. T. Kung**, A regular layout for parallel adders, *IEEE Trans. Comput.* **31** (3), 260–264 (1982).
10. **T. Han** and **D. A. Carlson**, Fast area-efficient VLSI adders, in *Proc. IEEE 8th Symp. Computer Arithmetic, Como, Italy, May 19–21, 1987* (IEEE, Los Alamitos, 1987), pp. 49–56.
11. **M. Snir**, Depth-size trade-offs for parallel prefix computation, *J. Algorithms* **4**, 185–201 (1986).
12. **J. Bund**, **C. Lenzen**, and **M. Medina**, Optimal metastability-containing sorting via parallel prefix computation, *IEEE Trans. Comput.* **69** (2), 198–211 (2020).
13. **Y.-C. Lin** and **L.-L. Hung**, Straightforward construction of depth-size optimal, parallel prefix circuits with fan-out 2, *ACM Trans. Des. Autom. Electron. Syst.* **14** (1), ID 15 (2009).
14. **M. Sheeran** and **I. Parberry**, A new approach to the design of optimal parallel prefix circuits, *Tech. Rep. No. 2006:1* (Chalmers Univ. Tech., Göteborg, 2006).

Igor S. Sergeev

Received December 22, 2023

Revised April 14, 2024

Accepted June 22, 2024