ПРОГРАММНАЯ РЕАЛИЗАЦИЯ ОДНОГО АЛГОРИТМА ВЫБОРА РАЗМЕРОВ ЭЛЕМЕНТОВ ИНТЕГРАЛЬНЫХ СХЕМ

П.Н. Розенштейн, Б.Д. Розенштейн, А.М. Камаев, И.Б. Мееров

Нижегородский госуниверситет им. Н.И. Лобачевского

Введение

Проектирование сверхбольших интегральных схем (СБИС) – сложный многостадийный процесс, значительная часть этапов которого требует решения оптимизационных задач в различных постановках для улучшения свойств схемы в смысле некоторых критериев, среди которых – уменьшение площади, занимаемой элементами интегральной схемы. Один из способов решения данной задачи – выбор размеров элементов (сайзинг).

Анализ литературы показывает, что за последние десятилетия эволюционировали не только способы решения, но и сами постановки задач: в рамках задачи сайзинга предлагалось оптимизировать не только площадь схемы, но и другие величины, на которые непосредственно влияют размеры элементов [2, 5-7], рассматривался сайзинг не только элементов, но и проводов, и т.д. Независимо от этого, исходная задача формулируется в виде задачи дискретной оптимизации и является NP-трудной [1]. Попытки решения такой задачи сравнительно редки (см., например, [1]). На практике чаще всего рассматривается непрерывное расширение, сформулированное в [5] в виде задачи выпуклого математического программирования.

В рамках данной работы решалась задача минимизации площади, занимаемой элементами схемы, при ограничении на максимальную задержку при прохождении сигнала. Решение указанной задачи позволяет не только уменьшать размеры интегральной схемы, но и повышать ее производительность. Задача решалась методом Лагранжа (Lagrangian relaxation), подробно рассмотренным в работах [2, 7] применительно к задаче сайзинга. В отличие от [2], в данной работе размеры сегментов проводов считаются постоянными, но при этом выполнено обобщение на широкий спектр возможных типов элементов схемы (в [2] рассматривают только сумматоры). Ряд алгоритмических идей, указанных в [2, 7], оказал существенную помощь при создании эффективной программной реализации.

Программная реализация выполнена в рамках исследовательского программного комплекса itlPlace [9], разрабатываемого на факультете ВМК ННГУ.

1. Постановка задачи

Рассмотрим схему, состоящую из K элементов. Пронумеруем их в топологически обратном порядке (то есть от выходных контактов схемы до входных контактов). Множество всех контактов схемы обозначим как $Pins$, множество выходных контактов схемы – $PrimaryOutputs$, множество входных контактов схемы – $PrimaryInputs$.

Будем рассматривать модель схемы, в рамках которой все элементы имеют одинаковую ширину y_k и разную длину x_k. Ассоциируем переменную a_{k_i} с контактом i элемента k. Переменная a_{k_i} обозначает время прибытия сигнала на контакт i элемента k (максимальную задержку от входа схемы до контакта). Тогда задача минимизации общей площади при условии ограничения на максимальную задержку (прямая задача $\mathcal{P}$) может быть сформулирована как

$\mathcal{P}: \sum_{k=1}^{K} x_k \rightarrow min$

при ограничениях:

1) для всех выходных контактов схемы i:

$$a_i \leq A_0, \tag{1}$$

где a_i – время прибытия сигнала на контакт i, A_0 – максимальная допустимая задержка сигнала на схеме.

2) для всех пар непосредственно соединенных между собой контактов i и j:

$$a_i + D_{ij} \leq a_j, \tag{2}$$

где D_{ij} – задержка сигнала, проходящего от контакта i до контакта j; для моделирования задержки D_{ij} используется формула Элмора [4]. Для входных контактов схемы w положим $a_w = 0$. Для удобства проведения выкладок учтем этот факт только после проведения преобразований.

3) ограничения на длину элемента (L_k и U_k – минимальная и максимальная длина элемента k):

$$L_k \leq x_k \leq U_k,\ k = 1, \cdots, K \tag{3}$$

Число ограничений в $\mathcal{P}$ полиномиально зависит от K и m. Для задачи $\mathcal{P}$ целевая функция – поkином (это было показано в [3]) и ограничения могут быть записаны в виде полиномов. Известно, что такая задача выпукла, то есть $\mathcal{P}$ имеет единственный минимум (глобальный минимум). Далее будем рассматривать эту формулировку задачи сайзинга.

2. Метод решения

2.1. Формулировка задачи Лагранжа $\mathcal{LR}/\lambda$

В соответствии с методом Лагранжа, для каждого ограничения на время прибытия сигнала вводятся неотрицательные переменные λ (множители Лагранжа). Для всех ограничений типа (1) вводим λ_{i0} (i – выходной контакт схемы). Для неравенств типа (2) – λ_{ij} (сигнал из контакта i поступает непосредственно на контакт j). Таким образом, для каждой пары соединенных контактов введен множитель Лагранжа.

Пусть λ является вектором всех введенных множителей Лагранжа, $\mathrm{x} = (x_1, \cdots, x_K)$ и a – вектор, компоненты которого времена прибытия сигнала на соответствующий контакт.

Пусть

$$L_\lambda(\mathrm{x}, \mathrm{a}) = \sum_{k=1}^{K} x_k + \sum_{i \in PrimaryOutputs} \lambda_{i0}(a_i - A_0) + \sum_{j \in Pins \backslash PrimaryInputs} \ \sum_{i \in input(j)} \lambda_{ij}\left(a_i + D_{ij} - a_j\right),$$

где $PrimaryOutputs$ – множество всех выходных контактов схемы, $Pins \backslash PrimaryInputs$ – множество всех контактов, кроме входных контактов схемы. Запись $j \in input(i)$ обозначает, что сигнал из контакта j непосредственно приходит в контакт i.

Тогда задача Лагранжа выглядит следующим образом:

$\mathcal{LR}/\lambda$: $L_\lambda(\mathrm{x}, \mathrm{a}) \rightarrow min$

при $L_k \leq x_k \leq U_k,\ k = 1, \cdots, K$.

Пусть функция $Q(\lambda)$ является оптимальным значением задачи $\mathcal{LR}/\lambda$. Составим двойственную задачу Лагранжа $\mathcal{D}$:

$\mathcal{D}$: $Q(\lambda) \rightarrow max$ при $\lambda \geq 0$.

Так как задача $\mathcal{P}$ является выпуклой, то, если λ является оптимальным решением $\mathcal{D}$, то оптимальное решение $\mathcal{LR}/\lambda$ будет также оптимумом для $\mathcal{P}$ (см. [8], теорема 2.13).

2.2. Упрощение задачи $\mathcal{LR}/\lambda$

В задаче $\mathcal{LR}/\lambda$ имеется два набора переменных: вектор $\mathrm{x} = (x_1, \cdots, x_K)$ и вектор a, компоненты которого – времена прибытия сигнала на соответствующий контакт. На переменные a не накладываются ограничения. Следовательно, в точке экстремума должно выполняться необходимое условие экстремума Куна-Таккера и $\frac{\partial L_\lambda(\mathrm{x,a})}{\partial \mathrm{a}} = 0$. С использованием этого утверждения после проведения ряда тождественных преобразований получаем некоторые условия оптимальности для λ. В результате при любых $\lambda \in \Omega_\lambda = \{\lambda \geq 0\colon \lambda$ удовлетворяет условиям оптимальности$\}$ задача $\mathcal{LR}/\lambda$ может быть переформулирована с существенным сокращением числа неизвестных:

$\mathcal{LR}/\lambda$: $\quad \mathcal{L}_\lambda(\mathrm{x}) = \sum_{k=1}^{K} x_k + \sum_{j\in Pins\setminus PrimaryInputs} \sum_{i\in input(j)} \lambda_{ij} D_{ij} \to min$

при $L_k \leq x_k \leq U_k,\ k = 1, \cdots, K$.

Для решения задачи $\mathcal{LR}/\lambda$ при фиксированном $\lambda \in \Omega_\lambda$ в работе используется высокопроизводительный решатель TAO (Toolkit for Advanced Optimization), разработанный в Argonne National Laboratory. Применяется метод переменной метрики с ограниченной памятью (limited memory variable metric method).

После нахождения оптимальных значений x, оптимальные a могут быть вычислены, если по очереди рассматривать переменные a_i в топологическом порядке и для каждого a_i устанавливать наименьшее возможное значение, которое удовлетворяет ограничениям $\mathcal{P}$.

2.3. Решение двойственной задачи $\mathcal{D}$

Как было показано выше, вместо всех $\lambda \geq 0$ можно рассматривать только такие, что $\lambda \in \Omega_\lambda$. Итак задача $\mathcal{D}$ может быть переформулирована следующим образом:

$\mathcal{D}$: $\qquad Q(\lambda) \to max$ при $\lambda \in \Omega_\lambda$,

где $Q(\lambda)$ является оптимальным значением для $\mathcal{LR}/\lambda$.

Функция $Q(\lambda)$ является вогнутой функцией от $\lambda \geq 0$ (см. [8], стр. 124). Однако в общем случае $Q(\lambda)$ не дифференцируема, это не дает применять методы, требующие вычисления градиента. Вместо этого обычно используется методы субградиентной оптимизации. Начиная из произвольной точки λ, метод итеративно перемещается из текущей точки в следующую точку, следуя направлению «субградиента». На шаге h сначала решается $\mathcal{LR}/\lambda$. Далее для каждого ограничения определяется «субградиент» как разность правой и левой частей ограничения. Новое значение λ получается как сумма прежнего λ и произведения направления «субградиента» на шаг ρ_h. После каждого такого перемещения λ проецируется на ближайшую точку в Ω_λ, для того чтобы можно было решать задачу $\mathcal{LR}/\lambda$ в упрощенном виде. Процедура повторяется до тех пор, пока метод не сойдется.

Известно (см., например, [10], стр.121), что если последовательность длин шагов $\{\rho_h\}$ удовлетворяет условиям $lim_{h\to\infty}\rho_h = 0$ и $\sum_{h=1}^{\infty} \rho_h = \infty$, то метод субградиентной оптимизации сойдется к оптимальному решению.

3. Алгоритм решения задачи сайзинга

Вход: исходный вектор x размеров элементов, начальный вектор множителей Лагранжа λ.

Выход: оптимальный вектор x размеров элементов.

1. Решить задачу Лагранжа $LR/(\lambda)$ и получить оптимальный вектор x при фиксированном λ. В данной работе используется решатель TAO.
2. Вычислить переменные a_i, присваивая им минимальные значения, удовлетворяющие ограничениям прямой задачи.

3. Сделать шаг решения двойственной задачи Лагранжа, обновив вектор λ при фиксированном x. Спроецировать новый λ на множество допустимых значений.
4. Для получения оптимального λ повторить шаги 1 и 2 до тех пор, пока не выполнится условие $(\sum_{k+1}^{K} x_k - Q(\boldsymbol{\lambda})) \leq \varepsilon$.
5. Повторить шаг 1 при найденном λ. Полученный вектор x является решением задачи.

4. Результаты экспериментов

Для тестирования реализации проведен ряд экспериментов на тестовых схемах из пакета IWLS 2005 Benchmarks Set (http://www.iwls.org/iwls2005/benchmarks.html). В качестве исходных использовались размещения, полученные инструментом itlPlace [9]. При оценке результатов, наряду с изменением площади схемы, рассматривались изменения интегральных временных характеристик схемы: TNS – Total Negative Slack (суммарная задержка сигнала на всех соединениях схемы) и WNS – Worst Negative Slack (максимальная задержка сигнала при прохождении от одного из входов до одного из выходов схемы).

Полученный в результате работы алгоритма сайзинга вектор размеров элементов отображался в ближайший вектор из множества существующих в библиотеке элементов с дискретными размерами. Поэтому имеет смысл рассматривать отдельно результаты работы алгоритма и результаты, полученные после округления.

В табл. 1 даны результаты экспериментов без учета последующего округления размеров элементов. Результаты позволяют сделать следующие выводы:

- ограничение на максимально допустимое время прихода сигнала не только не приводит к ухудшению, но и, напротив, приводит к некоторому улучшению временных характеристик схемы;
- зависимость процента относительного уменьшения площади носит обратно пропорциональный характер по отношению к проценту относительного улучшения временных характеристик.

Таблица 1. Результаты работы алгоритма сайзинга. Относительное изменение характеристик размещения

Интегральная схема	Число элементов	Относительное изменение TNS (%)	Относительное изменение WNS (%)	Относительное изменение площади (%)
IWLS_GP_002017_b12	2017	4,96	0,81	8,36
IWLS_GP_003389_wb_dma	3389	7,28	1,36	3,86
IWLS_GP_006724_s38584	6724	5,25	7,95	1,14
IWLS_GP_007959_systemcaes	7959	1,61	4,31	1,03
IWLS_GP_012458_b15_1	12458	5,60	1,94	1,79
IWLS_GP_020795_aes_core	20795	9,87	3,52	1,15
IWLS_GP_037117_b17	37117	7,45	0,87	0,64

В табл. 2 приведены изменения характеристик схемы с учетом отображения полученных размеров элементов на библиотеку. Данные результаты показывают, что

- Округление приводит к еще большему уменьшению площади.
- При уменьшении площади в результате округления ухудшаются временные характеристики. Однако это не всегда приводит к ухудшению относительно начальных TNS и WNS, и в ряде случаев исходная схема оптимизируется по всем трем параметрам.

Таблица 2. Результаты работы алгоритма сайзинга. Относительное изменение характеристик размещения после округления полученных размеров элементов

Интегральная схема	Число	Относительное	Относительное	Относительное

	элементов	изменение TNS (%)	изменение WNS (%)	изменение площади (%)
IWLS_GP_002017_b12	2017	3,50	-5,48	8,89
IWLS_GP_003389_wb_dma	3389	5,23	-4,04	3,92
IWLS_GP_006724_s38584	6724	4,88	5,88	1,35
IWLS_GP_007959_systemcaes	7959	0,48	2,12	1,12
IWLS_GP_012458_b15_1	12458	4,27	-3,37	2,04
IWLS_GP_020795_aes_core	20795	8,30	-1,35	2,20
IWLS_GP_037117_b17	37117	7,00	-0,09	0,99

Заключение

В рамках работы получены следующие результаты:

1. Проведен анализ литературы, описывающей задачу оптимального выбора размеров элементов СБИС и методы ее решения. По результатам анализа выбрана математическая постановка задачи и подход к решению, основанный на применении метода Лагранжа.
2. Выполнен ряд преобразований, позволивших упростить процесс поиска решения исходной задачи. Использованы идеи, опубликованные в печати и опробованные ранее на другом классе задач.
3. Разработан прототип программного обеспечения, решающий задачу выбора размеров элементов СБИС в указанной постановке.
4. Разработанное ПО интегрировано в исследовательский программный комплекс itlPlace.

В дальнейшем планируется расширение функциональности разработанного ПО.

Работа выполнена в лаборатории «Информационные технологии» ВМК ННГУ при поддержке компании Интел.

Литература

1. Chan P.K. Algorithms for library-specific sizing of combinational logic // Proc. ACM/IEEE Design Automation Conf., 1990. P. 353-356.
2. Chen C.P., Chu C.C.N., Wong D.F. Fast and exact simultaneous gate and wire sizing by lagrangian relaxation // IEEE Transactions on Computer-Aided Design, 1999. P. 1014–1025.
3. Duffin R.J., Peterson E.L., Zener C. Geometric Programming–Theory and Application. New York, Wiley, 1967.
4. Elmore W.C. The Transient Analysis of Damped Linear Networks with Particular Regard to Wideband Amplifiers // J. Applied Physics. 1948. Vol. 19(1).
5. Fishburn J., Dunlop A. TILOS: A posynomial programming approach to transistor sizing // ICCAD. 1985. P. 326–328.
6. Nikoubin T., Bahrebar P., Pouri S., Navi K., Iravani V. Simple exact algorithm for transistor sizing of low-power high-speed arithmetic circuits // VLSI Design. 2010. P. 1.
7. Wang J., Das D., Zhou H. Gate sizing by Lagrangian relaxation revisited // IEEE/ACM International Conference on Computer-Aided Design, San Jose, CA, 2007.
8. Городецкий С.Ю., Гришагин В.А. Нелинейное программирование и многоэкстремальная оптимизация: Учебное пособие. Нижний Новгород: Изд-во ННГУ, 2007.
9. Живодеров А.В., Камаев А.М., Корняков К.В., Мееров И.Б. Инфраструктура для исследования некоторых алгоритмов автоматизированного проектирования СБИС // Вестник Нижегородского университета им. Н.И. Лобачевского. 2010. №4(1). С. 170-176.
10. Мину М. Математическое программирование. Теория и алгоритмы. М.: Наука, 1990.